

OBITUARY NOTICE

EDWARD LEE THORNDIKE, 1874-1949

Thorndike is dead. The news comes as a blow to psychologists the world over. That prodigy of human energy and of experimental ingenuity is gone. His death marks the end of a chapter in psychology.

In order to appreciate his major contribution, one must go back to 1898, the year of his first considerable publication, *Animal Intelligence*, first brought out in the *Psychological Monograph* series and later reprinted as a book. It was the hey-day of introspection, of the description of 'experience subjectively regarded', of 'structuralism' as Titchener called it. Even the psychology of animals dealt with 'psychic events'. In the same year the physiologist Bethe devoted 100 pages of Pflüger's *Archiv* to a classical discussion entitled 'Can we ascribe psychic qualities to ants and bees?' Thorndike professed—and if in the enthusiasm of his twenty-four years he overestimated the uniqueness of his method we can forgive him—to apply for the first time the canons of experimental science to the behaviour of animals. Most studies of animals, he claimed, had been made by people anxious to prove how marvellous they are, 'defending animals against the charge of having no power of rationality'. Even psychologists had 'looked for the intelligent and unusual and neglected the stupid and normal'. Only secondarily does the monograph of 1898 ask the question, 'What do animals feel?'. One feels already the shadow of Behaviourism approaching. The experimental apparatus employed in the investigation was of course the puzzle box.

Thorndike's general presuppositions are worth noting. Measurement was to be made. That was part of the scientific programme. The experiment was conceived as one in motor association. It was repeated until the animal had formed 'a perfect association between the sense impression... and the impulse leading to the successful movement'.

Thus the concept of association is already fundamental for Thorndike's thinking, and in fact at the end of his monograph he states explicitly that he has 'made the beginning of an exact estimate of just what associations... an animal can form'. He has also rejected such 'higher' processes as reason, comparison or inference. And, most important of all, 'animal intellection is made up of a lot of specific *connections*' (italics by present writer); Columbia psychologists and through them American teachers were to hear about *connections* for the next fifty years.

In this way was born that rather chilling changeling, the 'trial and error' animal, which blunders about by instinct until it blunders on the correct course of action, and whose successful actions are 'stamped in' by 'the resulting pleasure'. The study was probably, as Thorndike claimed at the time, the first extended research on association in animals, at least of a quantitative nature; although a hedonistic-associational theory had, of course, previously been advocated by a number of people. Its results seem to have stamped Thorndike's thinking permanently.

Yet five years later, in 1903, quite characteristically appears a book entitled *Educational Psychology* but with no mention of the 'laws of learning', and in fact with no entry under 'learning' in the index. It 'attempts to apply to... educational problems the methods of exact science'. But in *The Elements of Psychology* (1905) the 'laws of frequency and

'recency' are definitely laid down, together with some others. The battle was joined. It has continued for forty-five years. The opponents have been sometimes other people, psychological theorists, experimentalists, educationalists, sometimes Thorndike himself. During the fray the *law of exercise* has almost disappeared as the result of Thorndike's own experiments. The *law of effect* has been modified and vigorously attacked; with some qualifications it was apparently held by Thorndike to the end of his life. Forty-nine years after Thorndike announced that the experiments on cats in a puzzle box showed that 'the particular impulse leading to the successful act will be stamped in by the resulting pleasure', the Psychological Bulletin devoted seventy-four pages to a discussion of the *Law of Effect*; included was a bibliography of 332 titles. The paper was followed by others, for and against. The ultimate issue has long been debated by moralists, criminologists and even metaphysicians. But it was Thorndike who first put the problems of repetition and of the effect of an action into experimentally verifiable form... and thereby revealed many difficulties that nobody had ever thought about before.

Thorndike's fundamental work on learning will probably be his monument for many years to come. One may stand in admiration of it without agreeing with its conclusions. It was prosecuted with daemonic vigour, great experimental fertility, and complete willingness to follow whither the experimental argument led. His name will endure in the list of the great Associationists which Hartley heads.

But his endeavours were not confined to the general theory of learning. Throughout the field of human education he made contributions. One thinks, almost at random, of his pioneer work on adult learning (1927), which shows *par excellence* his fertility of scientific invention, on handwriting (1910), mentioned by Thurstone in a recent discussion of the psychophysical methods, on the Psychology of Arithmetic (1922), on the effect of discussion on group decisions (1938), on the Measurement of Intelligence (1926). One remembers *The Teacher's Wordbook of 30,000 words* (1944), in which Thorndike, together with Lorge, determined the incidence of words selected from 18,000,000 in 285 sources; his classic study of Transfer (1924), his 'scale for the general merit of children's drawings' (1923), his *Study of Twins* (1905), his *Psychology of Wants, Interests and Attitudes* (1935), his massive *Man and His Works* (1943). These are but a random sample. In the last number of the Directory of the American Psychological Association he listed as his subjects of research: learning, tests and measurement, psychology of economics, government and law, psychology of language. It is a long list, but it does not begin to cover the activities of that prodigiously productive life.

When Thorndike was thirty-one years old William James said of him: 'Professor Thorndike has more horror of vagueness, of scholastic phrases, of scientific humbug than any psychologist with whom I am acquainted... (his) superabounding fertility in familiar illustrations of what he is describing amounts to genius.'

Truly a great psychologist has passed from among us.

GEORGE HUMPHREY

A STUDY IN THE RECOGNITION OF FIGURES¹

By J. C. TSAO, *From the University of Hong Kong*

I. *Introduction* (p. 57). II. *Method* (pp. 57-59). III. *Results* (pp. 59-64). IV. *Discussion* (pp. 64-66). V. *Summary and conclusions* (p. 67). *References* (p. 67).

I. INTRODUCTION

In studies on memory for figures, the reproduction method is generally used. In reproductions, which are, as noticed by Tennies (1942), much conditioned by the subject's drawing ability, various changes are bound to occur. These changes may be not entirely due to the dynamic changes in memory; for Hanawalt (1937) found that levelling and sharpening took place in direct copies of stimulus-figures made by subjects while the figures were being exposed. Hanawalt observed, furthermore, that in recognition the subjects chose the correct or approximately correct figures more often than they could draw them in reproductions.

Zangwill (1937) and Goldmeier (1941), as well as Hanawalt, have adopted the recognition method. But they used recognition only for a comparative purpose, and made no systematic study of the changes in recognition of figures.

The study reported in this paper was an attempt to investigate the changes in recognition of figures.

II. METHOD

(1) *Material*

In the present study, three experiments have been carried out, which differed from one another both in material and in procedure. In Exp. 1, twelve figures were used; most of them were well balanced and symmetrical. Each figure had ten forms. The difference between every two adjacent forms was slight but quite distinguishable. Of each figure, form no. 1 was the simplest or the smallest, form no. 10 the most complex or the largest, and form no. 5 or 6 was the intermediate one.

In Exp. 2, seven figures used in Exp. 1 were retained, but two additional forms were added to each of them, one at each end. Each figure had, therefore, twelve forms instead of ten. Five new figures, with twelve forms each, were added. Here too, therefore, there were twelve figures in all.

The twelve figures used in Exp. 3 were the same as in Exp. 2, but the number of forms in each figure was increased to twenty-one, so that the difference between every two adjacent forms was considerably slighter.

All forms of the figures were drawn each on a white card of 6×4 in. The size of each form was about 6-9 sq.in. In Fig. 1, some specimens of the material are presented.

¹ This study was carried out in the Cambridge Psychological Laboratory. The writer is much indebted to Mr G. C. Grindley for suggestions and help.

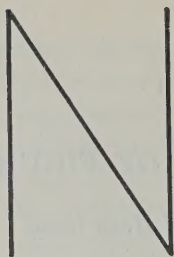
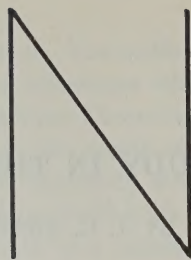


Fig. G. Form no. 5



Form no. 6

(Used in Exps. 1 and 2)

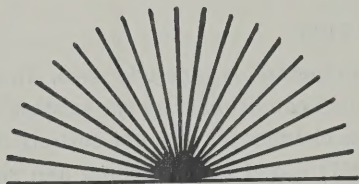


Fig. J. Form no. 6



Form no. 7

(Used in Exps. 1 and 2)

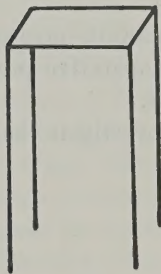
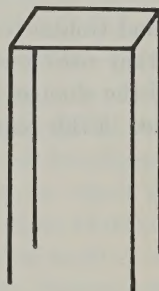


Fig. D. Form no. 15



Form no. 16

(Used in Exp. 3)

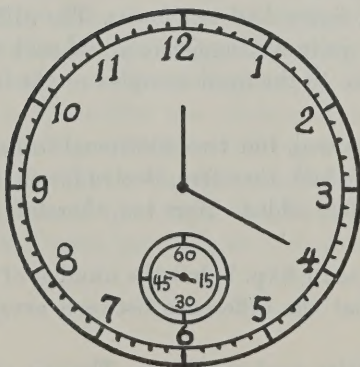
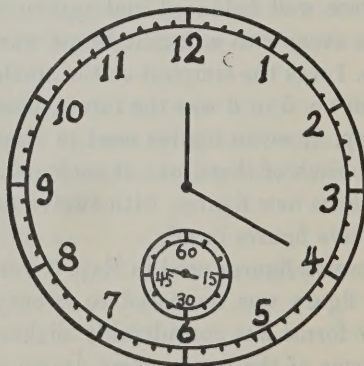


Fig. L. Form no. 14



Form no. 15

(Used in Exp. 3)

Fig. 1. Some specimens of the material.

(2) *Subjects*

Altogether 119 subjects participated in the experiments. Forty-eight subjects took part in Exp. 1, of whom three were members of the Cambridge Psychological Laboratory, eighteen were R.A.F. staff and twenty-seven were R.N. personnel. Forty-four took part in Exp. 2 and twenty-seven in Exp. 3, all of them naval personnel. None of the subjects had any previous experience in experiments on memory for visual forms, or knew the purpose of the present study before or during the experiments.

(3) *Procedure*

In Exp. 1, one form of a figure was exposed to the subject for 2 sec. After a certain interval, the ten forms of the figure were shown to the subject one by one, each for 2 sec. The subject had to identify the original form which was exposed before the interval by responding 'yes' or 'no' to the form presented. The particular form among the ten of a figure to be shown as the 'original' and the presentation of the ten forms in recognition were arranged in chance order. The forms of a figure used as originals and those used in recognition were in two duplicate sets. The different intervals were 0, 1, and 5 min. and 24 hr. After each interval, the recognition of three figures was tested.

In Exp. 2, during recognition, all the twelve forms of a figure were put on a table in front of the subject, who was asked to pick out the one which he considered to be the original. Besides the one first chosen, the subject had to select one or more forms, as alternatives, which he thought, next to the first one, likely to be the original. The intervals were 1 min., 10 min., and 24 hr. Each subject was tested for the recognition of four figures after each interval. In other respects the procedure was the same as that in Exp. 1.

The procedure in Exp. 3 was similar to that in Exp. 2, except in the following points: in recognition, the twenty-one forms of the figure were put in front of the subject who was asked first to indicate a range, or a certain number of forms, which would certainly include the original, then to select one form from the range as the original. The subject was instructed that the range of forms which he considered to include the original should be as narrow as possible. The intervals were 1 and 10 min., 24 and 72 hr. Each subject was tested for the recognition of three figures after each interval. As in Exps. 1 and 2, the twelve figures were used in rotation between different intervals and different subjects.

III. RESULTS

(1) *Accuracy of recognition at various intervals*

Tables 1-3 show the frequency distributions of the 'scores' of recognition in the three experiments. Score 0 means correct recognition; 1 indicates one step from the original towards form no. 10; -1, one step in the direction of form no. 1; 2 or -2, two steps mistaken in the one or the other direction; and so on. For example, if the stimulus-figure was form no. 6 of a certain figure and the subject responded positively to the form no. 6, his score would be 0; if he accepted form no. 5 in recognition, the score would be -1; form no. 4, score -2; if he accepted no. 7, the score would be 1; no. 8, score 2; and so forth.

Since in Exp. 1 the subject was allowed to respond affirmatively to more than one form of the figure as the original, the 'number of cases' at each interval in Table 1, instead of

being 144 (forty-eight subjects with three figures for each), is actually over it. In Exps. 2 and 3, the subject chose only one form as the original, and the 'number of cases' in Tables 2 and 3 is in accordance with the number of subjects multiplied by the number of figures on which each subject was tested.

The number of cases having a score of 0 is different after different intervals in all the three experiments. As shown in Tables 1-3, correct recognitions in Exp. 1 are: 0 interval,

Table 1. *Frequency distribution of scores of recognition in Exp. 1*

Score	Intervals			
	0	1 min.	5 min.	24 hr.
9	.	1	.	.
8	.	1	.	2
7	.	1	.	2
6	1	2	.	2
5	2	2	.	3
4	3	2	3	4
3	3	2	5	8
2	12	7	14	20
1	45	33	40	43
0	107	102	77	71
-1	29	36	34	38
-2	2	9	8	14
-3	2	3	1	8
-4	3	2	.	6
-5	.	.	1	2
-6	1	.	1	1
-7	1	.	.	1
-8	.	.	.	.
-9	.	.	.	1
Total	211	203	184	226
Mean	0.20	0.16	0.11	0.15
S.D.	1.43	1.65	1.32	2.30

Table 2. *Frequency distribution of scores of recognition in Exp. 2*

Score	Intervals		
	0	10 min.	24 hr.
5	.	1	1
4	2	2	4
3	3	2	5
2	11	12	14
1	26	33	30
0	105	74	81
-1	25	40	26
-2	4	12	14
-3	.	.	1
Total	176	176	176
Mean	0.19	0.07	0.16
S.D.	0.96	1.18	1.32

107; 1 min. interval, 102; 5 min. interval, 77; 24 hr. interval, 71. Converted into percentages, based on the maximum number 144, they are: 74.3, 70.8, 53.5, and 49.3 respectively. In Exp. 2, correct recognitions are: 105 or 59.7% after a 1 min. interval, 74 or 42.05% after a 10 min. interval, and 81 or 46.0% after a 24 hr. interval. In Exp. 3, correct recognitions are: 38 or 46.9% after a 1 min. interval, 30 or 37.0% after a 10 min. interval, 27 or 33.3% after a 24 hr. interval, and 18 or 22.2% after a 72 hr. interval.

With only one exception (10 min. and 24 hr. intervals in Exp. 2), the results of the

three experiments show that, the longer the interval between presentation and recognition, the less the correct recognition.

(2) *Changes in recognition and forms of figure*

Fig. 2 shows in graphic form the frequency distribution of the scores of recognition in each of the ten forms of the ten figures used in Exp. 1. Because both forms nos. 1 and 10 were extreme ones, the deviations in the recognition of them could only be in one direction. To a lesser degree this is also true of forms nos. 2 and 9. In forms nos. 5, 6, 4 and 7, the deviations are fairly symmetrical. The same trend was shown in the results of Exps. 2 and 3, though it was not so conspicuous because comparatively more forms and fewer

Table 3. *Frequency distribution of scores of recognition in Exp. 3*

Score	Intervals			
	1 min.	10 min.	24 hr.	72 hr.
13	.	.	1	.
12	.	.	.	.
11	.	.	.	1
10	.	.	.	1
9	.	.	.	.
8	1	.	2	.
7	.	1	3	.
6	1	3	3	2
5	.	.	1	1
4	2	4	3	6
3	2	1	1	9
2	9	3	7	8
1	16	10	14	10
0	38	30	27	18
-1	4	13	10	11
-2	7	13	4	8
-3	.	.	1	.
-4	.	1	2	1
-5	1	.	1	.
-6	.	1	1	1
-7	.	.	.	.
-8	.	.	.	1
-9	.	1	.	.
-10	.	.	.	2
-11	.	.	.	.
-12	.	.	.	.
-13	.	.	.	.
-14	.	.	.	.
-15	.	.	.	.
-16	.	.	.	.
-17	.	.	.	.
-18	.	.	.	1
Total	81	81	81	81
Mean	0.48	0.02	0.94	0.42
S.D.	1.72	2.35	3.00	3.86

subjects were used, and consequently the scores of recognition of each form were comparatively scanty.

(3) *Changes in recognition and time intervals*

All the frequency distributions after the various intervals in the three experiments, as shown in Tables 1-3, are fairly symmetrical; and all the mean scores are around zero, though slightly deviating towards the positive. These results show that the changes in recognition did not precipitate to one point, but diverged in every possible direction and rather followed the statistical rule of normal distribution.

The s.d., shown in Table 1, of the scores after the 24 hr. interval is the largest, while those of the scores after the other intervals do not differ very much from one another. The differences between the s.d. of the 24 hr. interval scores and each of those of the scores after the three shorter intervals are all statistically significant. This may indicate that while intervals under 5 min. do not differ much in their effects, 24 hr. between exposure

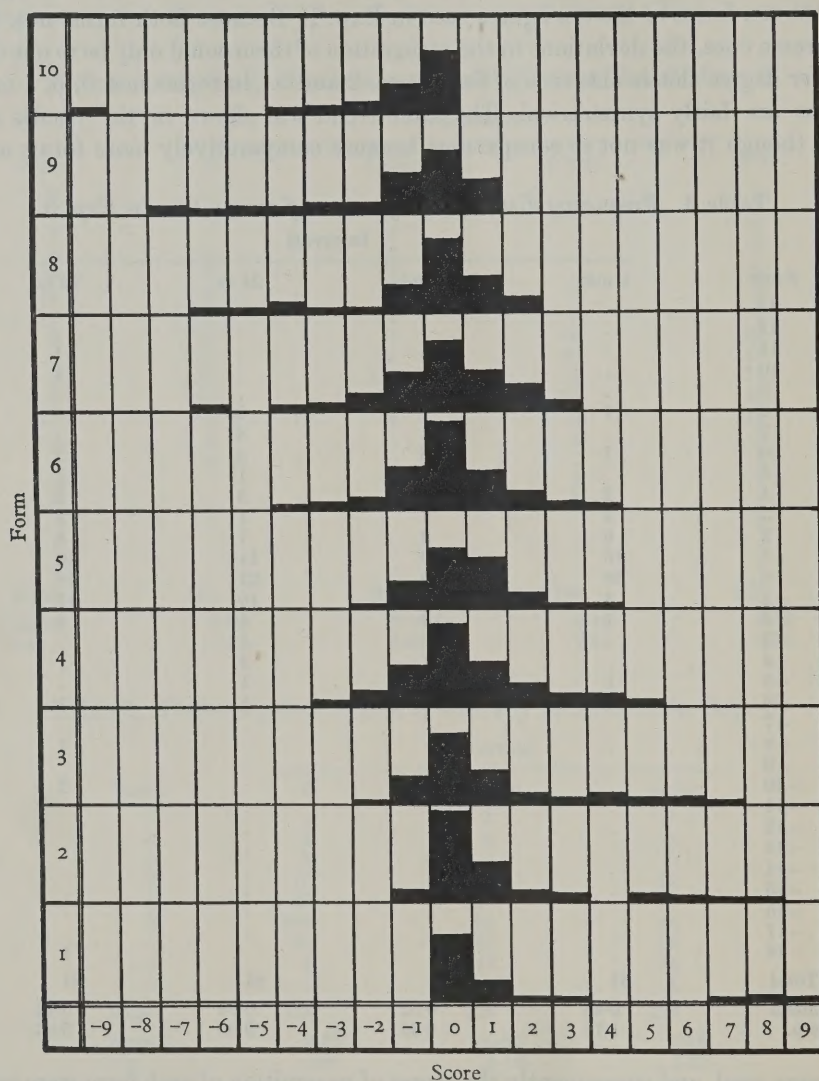


Fig. 2. Frequency histograms of the recognition scores of the ten forms of the figures used in Exp. 1.

and recognition cause the changes in recognition to become more widely scattered. The same trend is more remarkable in the results of Exps. 2 and 3, in which the s.d.'s of the scores of recognition increased with the length of the intervals between presentation and recognition. All the differences between the s.d. of the scores after a longer and after a shorter interval are statistically significant, since all the c.r.'s are around 3.00 or far

above it; there is only one exception which is the difference between the s.d. of the scores after the 10 min. interval and that after the 24 hr. interval in Exp. 2, but its c.r. is still 1.62. These results show clearly that, the longer the interval between presentation and recognition, the more scattered the changes in recognition.

In Exp. 2, as stated in the section on *procedure*, the subject, besides the first choice, took one or several forms as the 'possible' original stimulus; and in Exp. 3, he chose a range which, he thought, would include the original. The frequency distribution of the number of alternatives chosen by the subjects in Exp. 2 is presented in Table 4. The

Table 4. *Frequency distribution of alternative choices in Exp. 2*

No. of choices	No. of cases		
	1 min.	10 min.	24 hr.
0	26	20	9
1	82	74	80
2	54	66	61
3	12	15	18
4	1	.	6
5	1	.	2
6	.	.	.
7	.	.	.
8	.	1	.
Total	176	176	176
Mean	1.34	1.47	1.65
S.D.	0.86	0.94	0.93
Dif.	0.13	0.18	0.31*
C.R.	1.35	1.82	3.26

* Between 1 min. and 24 hr. means.

Table 5. *Frequency distribution of the number of figures chosen in Exp. 3*

No. of figures	No. of cases			
	1 min.	10 min.	24 hr.	72 hr.
1	9	1	2	.
2	39	22	12	6
3	29	40	31	23
4	3	16	22	22
5	1	2	12	20
6	.	.	2	6
7	.	.	.	2
8	.	.	.	1
9	.	.	.	1
Total	81	81	81	81
Mean	2.36	2.95	3.44	4.15
S.D.	0.768	0.786	1.067	1.347
Dif.	0.59	0.49	0.71	
C.R.	4.84	3.38	3.76	

frequency distribution of the number of forms within each range chosen by the subjects in Exp. 3 is given in Table 5. From Table 4 it may be seen that in Exp. 2 the subject had a clear tendency to choose more alternatives after longer intervals. The difference between the average number of alternatives chosen after 24 hr. and after 1 min. is highly reliable. The results of Exp. 3 show clearly that there was a progressive extension of the range chosen by the subjects as the intervals between presentation and recognition became longer. The differences between every two adjacent intervals are all significant. The number of forms, chosen as alternatives, or to be included in a range, may indicate the

accuracy in recognition; the larger the number, the less definite the recognition. The results of Exps. 2 and 3 show once more, therefore, that recognition became less and less definite as the interval between presentation and recognition lengthened.

IV. DISCUSSION

(1) *Influence of material*

The frequency histograms in Fig. 2 show how the form of figures has influenced the direction of changes in recognition. Figures with forms of the one extremity caused changes in the one direction, figures with forms of the other extremity yielded changes towards the other direction, and those with intermediate forms gave rise to changes in both directions. It sounds as if this were in accordance with Koffka's (1935) theory which assumes that 'the changes which any pattern undergoes must be determined by the pattern itself'. But the accordance breaks down if we go further with Koffka. He maintains that 'according to the nature of the pattern lines may gradually become straight or more curved, longer or shorter'. If this were true, only one type of change from a certain pattern would be possible, or at least would be dominant. Our results demonstrated, however, that in the intermediate forms, changes in recognition were in both directions. Neither of the changes towards the one or the other direction could be designated as the predominant one. Moreover, even the more or less unidirectional changes in the extreme forms did not exclude changes towards the opposite direction. Frequently figures of a rather complicated form gave rise to changes towards more complicated forms, and figures of very simple forms gave rise to changes towards still more simplified forms. Changes of this kind were already noticeable in the scores of recognition of the forms nos. 2 and 9, shown in Fig. 2. It was most remarkable in some cases in Exp. 3, when forms nos. 4, 5, 18 and 20 were used as originals. The distributions of the recognition scores of these four forms are as follows:

Form no. 4 as stimulus:								
Score	-2	-1	0	1	2	3	...	10
Frequency	3	2	6	2	.	1	.	1
Form no. 5 as stimulus:								
Score	-2	-1	0	1	2	3	4	5
Frequency	1	3	8	.	1	.	.	1
Form no. 18 as stimulus:								
Score	-1	0	1	2	3	.	.	.
Frequency	2	6	5	3	2	.	.	.
Form no. 20 as stimulus:								
Score	-18	...	-9	...	-2	-1	0	1
Frequency	1	.	1	.	3	3	12	7

It should be noted that forms nos. 4 and 5 were simpler forms, and more than half of the changes in recognition of them were towards still simpler forms (indicated by minus scores); and that nos. 18 and 20 were complex forms, and in the recognition of them the changes with the greatest frequency were one step towards greater complexity.

Changes towards extremity and changes towards intermediate forms are nothing new; they are called by Wulf (1922) sharpening and levelling. The fact that there are opposite changes does not, however, fit in quite well with the Gestalt theory according to which all changes should precipitate towards 'good figures'. Koffka cited Wulf's (1922) and Perkins' (1932) results as the main evidence of his theory. Both Wulf and Perkins used simple, non-symmetrical and irregular figures which naturally favour changes towards

symmetry and other standard forms. Had perfectly symmetrical figures been used by these authors, the changes would have been mainly towards less perfect forms.

Changes in memory for figures are influenced by the forms of the figures, precipitation to a certain good figure is, however, rather the exception than the rule.

(2) *Forgetting*

In all three experiments of the present study, the percentage of correct recognition decreased as the length of intervals between presentation and recognition increased. A decreased accuracy in recognition implies that forgetting has taken place. As shown in the results, forgetting was most pronounced in Exp. 3 and least in Exp. 1. This may be due to differences in the procedure. In Exp. 1, the number of forms to which the subject might respond positively was not limited, so that he had greater chance to get the correct one. In Exps. 2 and 3, only one form was allowed to be selected as the original, and there were more forms of each of the figures used in Exps. 2 and 3 than in Exp. 1, and in Exp. 3 than in Exp. 2; consequently the recognition might be most difficult in Exp. 3 and least in Exp. 1.

Formerly, Baldwin & Shaw (1895) found that accuracy in the recognition of squares of different size decreased as the interval between presentation and recognition lengthened. Piéron (1920) obtained an Ebbinghausian retention curve in learning and relearning figures of straight lines intersecting at various angles. Later investigators, using more complex figures than those of Baldwin & Shaw and Piéron, seldom mentioned forgetting in memory for figures. This may be due to the fact that nearly all of them used the reproduction method. Reproductions of figures are less subject to quantitative treatment, and generally in the reproduced figures the form of changes attracts the attention of the investigators more than the degree or magnitude of the deviations from the original. Luh (1922), and many others, have long proved that in rote learning recognition usually gives higher retention scores than reproduction, and Hanawalt's results indicated the same in memory for figures. If forgetting occurs in recognition, as is shown in the results of the present study, there should be forgetting in reproduction. In fact, the changes in reproduced figures are themselves indications of forgetting.

If the forgetting of figures were thoroughly studied, a comparison between retention and forgetting in verbal and rote learning and in memory for figures would be made possible. Such a comparison would bring together the results of studies on memory for figures on the one hand, and those obtained by investigators following the lead of Ebbinghaus on the other hand; so far, these two schools of studies have been proceeding along different lines and supporting different theories of memory.

(3) *Range of changes and time intervals*

The increases, in the magnitude of the s.d.'s of the scores of recognition in all the three experiments, in the number of alternative choices as the original stimuli in Exp. 2, and in the range of forms within which the original form was believed to be included in Exp. 3, all show that after longer intervals recognition was more diffuse. That recognition becomes less and less definite with the passage of time can be considered as an evidence of forgetting. But the magnitude of the range of positive responses to a certain stimulus has rarely been taken as a measure of retention and forgetting in former studies of memory; it has generally been used in studies on transposition and generalization, especially the latter.

The results of Exps. 2 and 3 of the present study showed that after a longer interval the range of forms eliciting positive responses increased. This resembles very much the generalization in conditioned responses (Hilgard & Marquis, 1940), though, in the studies of the latter, time interval is seldom the variable factor. Could the results of Exp. 2, and still more those of Exp. 3, be interpreted in this way to mean that a process of generalization had been taking place with the lapse of time so that after a longer interval the subject responded positively to more forms of the original figure? There is plenty of evidence in the psychology of learning which indicates that after a sufficient interval the response of a subject to a certain kind of stimulus generalizes, i.e. the range of relevant stimuli widens and the responses become more diverse. In memory for colours, Hanawalt & Post (1942) found 'a fairly wide area of indifference or equality; when the subject first reaches this area, he responds "equal"'. This 'fairly wide area' of remembered colours, just like the range of forms in Exp. 3 of the present study, on the one hand lacked definiteness as far as its stimulus was concerned; on the other hand, it was a generalized response to colours within a certain range of saturation.

In conditioned responses, generalization first occurs at the early stage of conditioning when the specific conditioned response has not been well established; it takes place also after extinction has begun (Hilgard & Marquis, 1940). Apparently, generalization is due to lack of differentiation, and it may occur either before the development of differentiation or when the differentiation is lost. In the present experiments generalization took place after certain intervals between presentation and recognition. Obviously the generalization was a result of the waning of differentiation which must be at its climax just after presentation. Had recognition been tested immediately after presentation under different conditions of exposure, such as longer and shorter exposure time or more complex and simpler forms, perhaps we could detect generalization before the development of differentiation.

In the present experiments the increases in the magnitude of s.d.'s, in the number of alternative choices, and in the range of forms took place synchronously with the decrease in correct recognition; in other words, generalization and forgetting occurred together. What is their relation? Is the one a part of, or an essential condition for, the other? Certainly there can be forgetting without generalization, as in the case where a particular word is just forgotten. There may be also generalization without forgetting, such as the generalization taking place before the establishment of a specific conditioned response. In the present experiments, however, the generalization was obviously due to forgetting; generalization took place because the subject could not make the exactly correct recognition. It seems that with the passage of time the memory for figures partly faded away, and as a reaction to forgetting certain processes were induced to enable the subject to meet the situation demanding retention. In recognition, the subject, to 'play safe', reacted positively to as many forms as necessary; this is generalization. In experiments where reproductions were required the subject tried to reconstruct the figure from what he remembered; this is reconstruction. Both generalization and reconstruction enabled the subject to meet the situation which demanded of him exact recognition or reproduction. It seems anthropomorphic to relate the generalization of conditioned reflexes in experiments on animals with the 'play safe' recognition in the present study, but there is little ground to consider these two cases of generalization as fundamentally different.

V. SUMMARY AND CONCLUSIONS

Three experiments on recognition of figures have been carried out; 119 subjects took part in them. Twelve figures with ten, twelve and twenty-one forms each were used in the first, second and third experiments respectively. One form of a figure was first presented to the subject for 2 sec., and recognition was tested with all the forms of the figure either immediately or after intervals ranging from 1 min. to 72 hr. In the first experiment, in recognition, the forms of a figure were shown one by one to the subject, who had to respond positively to the one or more forms which he thought to be the original. In the second experiment the forms of a figure were put in front of the subject and he was asked to select one as the original and several as alternatives. In the third experiment the forms were also shown *en masse* to the subject who first chose a range of forms within which the original was supposed to be included, and then selected one of the forms as the original.

The results showed that when intermediate forms were used as stimuli the deviations from the original in recognition spread according to the normal distribution; that when extreme forms were used the deviations were towards one direction since that was the only possibility; that as intervals between presentation and recognition lengthened, percentages of correct recognition decreased; and that as intervals became longer, the deviations in recognition were scattered more widely, and the subject tended to take more forms as the original stimulus. The results of the experiments led to the following conclusions:

(1) The changes in recognition of figures are influenced by the form of the figure, but they do not precipitate unidirectionally towards any 'good' figure, and their distribution is rather in agreement with the statistical rule of normal distribution.

(2) Forgetting occurs in memory for figures as well as in rote learning. Findings from studies on the retention of figures and of verbal material can, and should, be integrated in order to attempt a more complete understanding of memory.

(3) Generalization takes place in recognition of figures. This makes it possible to relate the results of studies on memory for figures and studies on conditioned reflexes, and thus to advance a step further towards a more comprehensive understanding of the learning processes in general.

REFERENCES

- BALDWIN, J. M. & SHAW, W. J. (1895). Memory for square size. *Psychol. Rev.* II, 236-9.
 GOLDMEIER, E. (1941). Progressive changes in memory traces. *Amer. J. Psychol.* LIV, 490-503.
 HANAWALT, N. G. (1937). Memory trace for figures in recall and recognition. *Arch. Psychol.* no. 216.
 HANAWALT, N. G. & POST, B. E. (1942). Memory trace for colour. *J. Exp. Psychol.* XXX, 216-27.
 HILGARD, E. R. & MARQUIS, D. G. (1940). *Conditioning and Learning*. New York.
 KOFFKA, K. (1935). *Principles of Gestalt Psychology*. London.
 LUH, C. W. (1922). The conditions of retention. *Psychol. Monog.* XXXI, no. 142.
 PERKINS, F. T. (1932). Symmetry in visual recall. *Amer. J. Psychol.* XLIV, 473-90.
 PIÉRON, H. (1920). Recherches comparatives sur la memoire des formes et celle des chiffres. *Année Psychol.* XXI, 119-48.
 TENNIES, L. G. (1942). Memory trace and perception in the blind. *J. Exp. Psychol.* XXX, 23-39.
 WULF, F. (1922). Beiträge zur Psychologie der Gestalt: VI. Über die Veränderung von Vorstellungen (Gedächtnis und Gestalt). *Psychol. Forsch.* I, 333-73.
 ZANGWILL, O. L. (1937). An investigation of the relationship between the processes of reproducing and recognizing simple figures, with special reference to Koffka's trace theory. *Brit. J. Psychol.* XXVII, 250-76.

(Manuscript received 25 July 1949)

THE DISTRIBUTION OF SCORES ON RAVEN'S PROGRESSIVE MATRICES TEST

BY BERNARD NOTCUTT, *From the University of Natal, Durban*

Mr Raven's article⁽²⁾ in this *Journal*, of September 1948, discussed among other topics, the changing form of the distribution of scores of Progressive Matrices at various ages. The norms given for the group test in the 1940 manual⁽¹⁾ show a population that is positively skewed at the age of 8 years, normally distributed at 9 years, and subsequently develops a marked negative skewing, which becomes increasingly prominent from 10 years onwards. An interesting comparison is provided with the scores obtained by Zulu subjects tested during 1948 in the neighbourhood of Durban. The test was being standardized for Zulus living under urban and peri-urban conditions, and the sample can be considered a good one from 11½ to 15½ years. Table 1 shows the distribution of scores obtained for this age range.

Table 1. *Progressive Matrices test: Scores of 762 Zulu primary school pupils from 11 years to 15 years 11 months*

Score	11 yr. +	12 yr. +	13 yr. +	14 yr. +	15 yr +
50-54	—	1	—	—	—
45-49	—	2	3	3	10
40-44	—	1	7	7	16
35-39	2	7	12	19	16
30-34	8	20	21	19	17
25-29	15	24	23	28	19
20-24	28	27	36	26	32
15-19	49	50	30	27	19
10-14	36	22	20	15	20
5-9	7	4	7	5	2
Total	145	158	159	149	151
Mean	18.3	22.0	23.7	24.1	26.9
S.D.	6.50	8.38	9.37	9.50	10.73

It is curious to note that over an age range where Raven's English sample shows a marked negative skewing, the Zulu sample shows an equally pronounced positive skewing. Fig. 1 is designed to emphasize this peculiarity. It represents the deviations from the median of scores obtained on the Progressive Matrices test of English and Zulu subjects at 11½, 12½ and 13½ years, for which comparable data are available. The English data are taken from the norms given in Raven⁽¹⁾.

If, instead of equating the ages of the two samples, we equate the median scores, we find that the distributions are remarkably similar. The median score on Progressive Matrices for the Zulu sample at 11½ years is almost exactly the same as that of the English sample at 8 years; similarly Zulu 12½ years equals English 8½ years; Zulu 13½ years equals English 9 years, and Zulu 14½ years is only a little behind English 9½ years. When deviations from the median score are compared for these groups as in Fig. 2 the patterns of distribution are seen to be similar. In each group there is a change from a positive skewing when the median score is 18 points, to a normal distribution when the median is about 25 points.

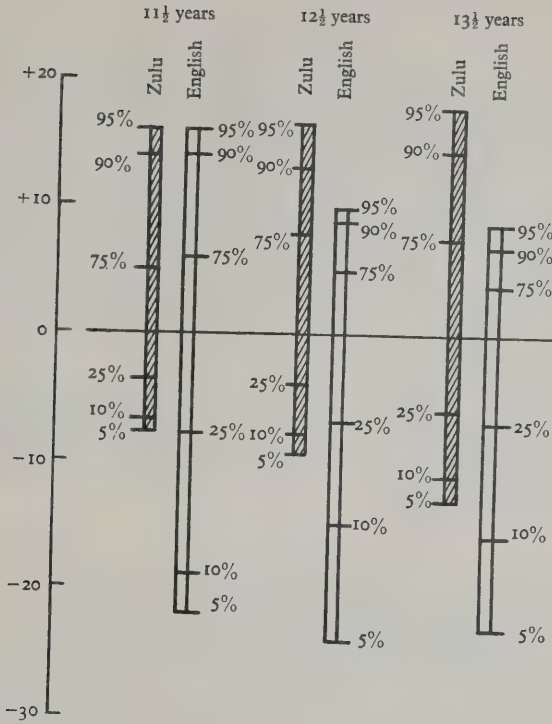


Fig. 1. Progressive matrices test: comparing deviations from the median in English and Zulu groups at three age levels.

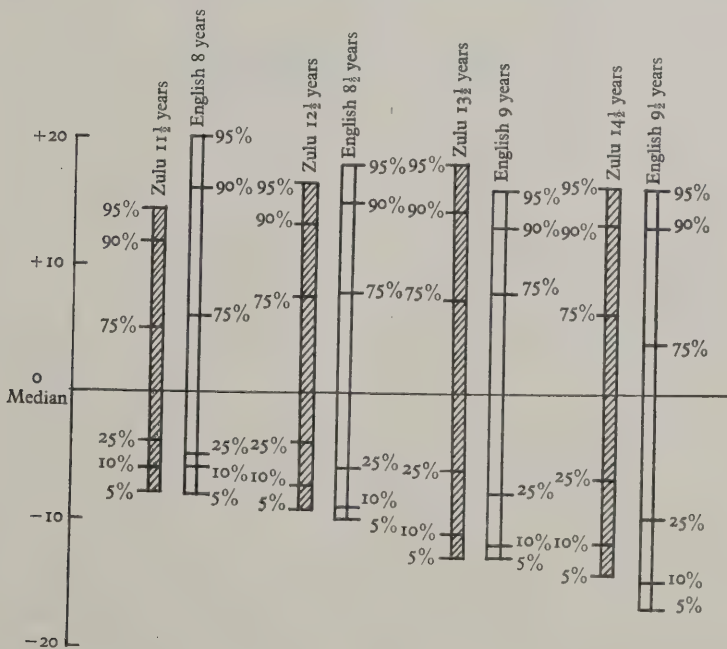


Fig. 2. Progressive matrices test: comparing deviations from the median in English and Zulu groups whose medians are approximately equal.

CONCLUSIONS

It is apparent from these comparisons that the form of distribution of an age-group sample of these scores depends more on the test than on the group tested. To increase one's score on Progressive Matrices from 44 to 50 is a far more difficult task than to increase it from 24 to 30, hence the negative skewing at higher levels. Thus the English norms are negatively skewed at ages where the Zulu norms (with a much lower average score) are positively skewed. This fits in with the view of Slater(3). A test like the Progressive Matrices is not really suitable for making inferences about the 'true' distribution or the 'real' rate of growth of intelligence. For studying this, a test built on the principles of Thorndike's(4) CAVD is required.

The data about Zulu scores on Progressive Matrices were obtained in the course of an investigation financed by the South African National Council for Social Research. The field-work was carried out by Mr S. B. Theunissen, assisted by Mr N. Mkele, who also did a large part of the computation.

REFERENCES

- (1) RAVEN, J. C. (1938). *Progressive Matrices*. London: H. K. Lewis.
- RAVEN, J. C. (1940). *Instructions, Key and Norms*. London: H. K. Lewis.
- (2) RAVEN, J. C. (1948). The comparative assessment of intellectual ability. *Brit. J. Psychol.* xxxix, 12-19.
- (3) SLATER, P. (1948). Comment on 'The comparative assessment of intellectual ability'. *Brit. J. Psychol.* xxxix, 20-21.
- (4) THORNDIKE, E. L. *et al.* (1926). *The Measurement of Intelligence*. New York: Teachers College, Columbia Univ.

(Manuscript received 30 May 1949)

A STUDY OF THE ISHIHARA TEST FOR COLOUR BLINDNESS¹

By R. W. PICKFORD, *From the Psychology Department, Glasgow University*

I. *Introduction* (pp. 71-72). II. *Comparison with the 'Rotating Disks' Test* (pp. 72-74). III. *Comparison with the 'Nine-Colour' Test*, (pp. 74-76). IV. *Comparison with the 'Four-Colour' Test* (pp. 76-78). V. *Conclusions* (pp. 78-79). VI. *Summary* (pp. 79-80). *References* (p. 80).

I. INTRODUCTION

The following three questions must be asked in a consideration of the adequacy of the Ishihara Test for Colour Blindness:

(1) does this test effectively separate major red-green defectives from the normal population;

(2) does it effectively separate different types of major red-green defectives from each other;

(3) can it be used to distinguish different degrees of defect among the normal population?

Collins(2) has given evidence that the Ishihara Test does show the presence of red-green deviations which can be detected with the anomaloscope, but in the Special Report of the Medical Research Council on Colour-Vision Requirements in the Royal Navy(5) it is indicated that serious discrepancies may be found between the results of different tests for red-green blind subjects. Vernon & Straker(10) have shown that the percentage of the male population of Great Britain found to be red-green defective on a combined form of the Ishihara and Stilling Tests varied from 5.37 to 9.45%, according to the district sampled. There was a statistically significant tendency for fewer defectives to be found in the north and east than in the south and west. These authors consider that the largeness of the average proportion of defectives was due to the test used, picking out many of the 'colour weak' and 'anomalous trichromats' in addition to the colour blind. Similar percentages, and interesting differences between groups, have been found by Geddes(3) and other investigators (2, 2a, 4). In the research, of which part is reported in this paper, 7.8% of major red-green defective men were found purely by chance(8).

In order to attempt to answer the three questions given above, a comparison was made between the 25-Plate Ishihara Test (Japanese printing) and certain other tests of red-green defects, in the course of a research reported elsewhere(8), and some of these results have appeared in *Nature*(7).

The first test with which the Ishihara Test was compared was a form of the Rayleigh Equation, using rotating disks made from Hering's coloured papers. This was similar to a test with rotating disks reported by Collins(1). It will be called the 'Rotating Disks' Test. The second test with which the Ishihara Test was compared was an anomaloscope test in which five pairs of opposite or complementary colours were studied, namely, red and green, orange and blue-green, yellow and blue, yellow-green and violet, and green and magenta. Since the green used in the red-green and in the green-magenta tests was the same, this will be called the 'Nine-Colour' Test. Only the red-green part of it will be dealt with here. The third test with which the Ishihara Test was compared was a slightly

¹Based on a paper read to the Psychology Section of the British Association, at Newcastle-upon-Tyne, 1949.

different form of this anomaloscope test, in which only four colours were used, namely, red and green, and yellow and blue. This will be called the 'Four-Colour' Test. In all tests any darkening of the red end of the spectrum was measured when it occurred.

Data arising from these three comparisons will be presented and discussed in this paper, and, since all three tests named above have been described elsewhere⁽⁸⁾ details of their techniques will not be given here.

II. COMPARISON WITH THE 'ROTATING DISKS' TEST

In the Rotating Disks Test many of the subjects were asked to do the Ishihara Test because it was hoped at first that the degree of colour weakness might be revealed by the number of errors made in the Ishihara, and that there might be a close relation between the two tests. This hope was altogether abandoned after a more careful study of the real nature of the problems involved, and a review of the results obtained will show how slender a hope it was.

In this and the other experiments reported in this paper the Ishihara Test was used in a dark room, and was illuminated with a 60 W. pearl lamp in a reflecting holder and fitted with a $\frac{1}{2}$ W.-to-daylight filter. This lamp was placed at a distance of about 1 m. from the test, and the subject read the test from about the same distance. The lamp was placed in such a position that the subject could not take advantage of the differential reflexion of light from the various coloured inks in the Ishihara Test. This is very important when using this test either in daylight or in artificial light. Many colour-blind subjects can read figures which 'shine' in a certain light, while normal subjects can read the alternative figures for colour-blind subjects often in the same way. The subject must be prevented from moving his head from side to side, or tilting the test, in order to take advantage of this differential reflexion while doing the test. The 'daylight' filter converts $\frac{1}{2}$ W. light to

Table 1. *Men in Rotating Disks Test: Ishihara result*

Subjects (type and no.)	Ishihara Test result			
	Correct	Multiple	Errors	No reading
Protanopes:				
Extreme 3	1, 2, 1	0, 0, 1	9, 15, 16	15, 8, 7
Moderate 3	1, 2, 2	2, 1, 3	11, 11, 13	11, 11, 7
Deutanopes:				
Extreme 3	1, 1, 2	2, 1, 1	15	7, 8, 7
Moderate 5	1	3-4	12-13	8-9
Green anomalous 2	2, 2	8, 2	9, 13	6, 8
R-G weak 13	Mode = 21	1	4	0
	Scatter = 14-24	0-7	0-4	0-1
Normal 9	Mode = 20	2	2	0
	Scatter = 14-24	0-8	1-4	0-5

Table 2. *Women in Rotating Disks Test: Ishihara result*

Subjects (type and no.)	Ishihara Test result			
	Correct	Multiple	Errors	No reading
Protanopes:				
Moderate 2	2, 1	5, 4	11, 12	7, 8
Deutanopes:				
Extreme 1	1	0	16	8
Moderate 1	2	6	11	6
R-G weak 18	Mode = 22	0	4	0
	Scatter = 15-25	0-7	0-7	0-1
Normal 27	Mode = 21	0	3	0
	Scatter = 16-25	0-6	0-6	0-1

normal white light, and is much more constant than even the light from a north window. The test plates were kept in the dark when not in use, and were thoroughly clean, and the subjects were tested throughout with great care. Tables 1 and 2 show the results for fourteen men and four women who were red-green blind, two men who were green anomalous, thirteen men and eighteen women who were red-green weak and nine men and twenty-seven women who were normal. The red-green weak are a division of the normal whose matching range (or double differential threshold) in the Rotating Disks Test was two or more times the modal range. A distinction was made between moderately and extremely red-green blind, in terms of matching range. This distinction and that between protanopes and deuteranopes, as well as the distinction between anomalous and colour blind, have been explained elsewhere (8).

In these tables the numbers of correct readings of single plates in the Ishihara Test are shown for the colour blind, and for the anomalous subjects individually. For the colour weak and normal subjects the modal frequencies of correct readings and their scatter are shown. The frequencies of multiple readings, unqualified errors and complete failures are given in similar ways. A perfect reading of the test requires 25 correct interpretations, including no reading of each of the four plates which are supposed to be illegible to the normal. A complete failure requires correct reading of the first plate, which is used to show the subject what he is expected to be able to do, eight plates with no reading, sixteen simple errors and the reading of four plates illegible to the normal. No double readings or alternatives should be given by anybody if the test is perfectly satisfactory, but in fact many normal subjects can read most of the figures which are supposed to be illegible to them, if they try to find alternatives.

The tables show that the Ishihara Test successfully failed all the colour-blind and anomalous subjects, even on a level so lenient as 2 correct readings in 25 (one of which is the first plate or 'joker'). The two green anomalous subjects were not distinguished at all by this test, and if this test were all we had to guide us, we should not suspect their existence as different from the colour blind. It is clear that the Ishihara Test would fail a number of normal and red-green weak subjects if the standard of failure were as strict as 14-18 correct readings, but the numbers of correct, alternative, faulty and blank readings gave absolutely no guide to the degree of a subject's weakness, if the major defectives are excluded. It is quite unknown whether any of the red-green weak subjects might be dangerous in practical situations, but, if they are dangerous and ought to be failed, then the Ishihara Test would be totally useless to discriminate them from the perfectly normal. If the Test were used with variable lighting, or if the plates were soiled (as I have been told they are sometimes where large numbers of recruits are tested), or if the testing were in the hands of those who have not had serious laboratory training, then it would be even less reliable. Indeed, variations in the quality of daylight from the north-east to the south-west parts of the country might account to some extent for the frequency differences reported by Vernon & Straker. In the British printing of the Ishihara Test, the composition of the light reflected by the inks is not always the same as that of the Japanese printing, though the hues appear to be almost the same, and this would be another possible source of error. This was mentioned by Riddell in a lecture, but not in its published summary (9).

A proportion of the red-green weak and normal subjects who made about the modal range of errors in the Ishihara Test were yellow-blue weak, and for a time it was thought

that yellow-blue weakness might have an effect on the results with this test. A careful tabulation of the results, however, showed no decisive relation between yellow-blue weakness and frequency of errors on the Ishihara Test. On the other hand, certain subjects who had very good colour vision were specially clever at picking out the figures they were not supposed to see in the Ishihara Test as well as the correct figures, and this was a frequent cause of as many as 4-8 'mistakes' and/or double readings. Some subjects with particularly good colour vision gave four or more alternative readings for a single plate. This shows imagination and quickness at spatial integration of dot patterns, and in some ways the Ishihara Test is like a projection test made with ink blots, in which one can find a variety of interpretations. Another source of errors was anxiety, especially where the subject was worried about his or her vision. The inability to converge the eyes owing to anxiety was a frequent symptom in one subject, and caused her to fail on the Ishihara Test, but it was clear that she had this difficulty often in ordinary reading, and had no colour weakness.

Vernon & Straker were probably correct in their supposition that the Ishihara-Stilling Test used in the Navy picked out many 'colour-weak or anomalous trichromats' as well as the strictly colour blind in the research they reported. Pseudo-isochromatic tables, however, are unreliable for detecting those with small weaknesses of colour sensitivity, because the number of errors made is no guide to the degree of colour weakness present. If anything, it might be said that the Ishihara Test is a worse guide than Stilling's Tables, because of the numerous alternative readings, which, though interesting, do not add to its value as a colour-vision test except in one respect, namely, that they give the major defectives something they are able to read, which maintains their confidence, while in Stilling's tables they would tend to pass page after page with complete inability to read anything at all. One extreme deuteranope reported that he was convinced that he was abnormal only by his ability to read in the Ishihara Test four figures which were illegible to the normal subject.

Before leaving this section it will be useful to comment on Vernon & Straker's terms 'colour-weak' and 'anomalous trichromat'. Their use of these terms might suggest that the colour weak and anomalous are closely similar, or perhaps that they are two degrees in a continuous series of variations: normal, colour weak, anomalous and colour blind just as we have arbitrary subdivisions in the continuous grading of intelligence quotients. In a careful study of more than 1000 individuals, however, the writer has shown that the variations of colour vision are strongly discontinuous and correspond to seven Mendelian allelomorphs: normal (which includes the deviant and colour weak with those who are absolutely normal), green anomalous, deviant deuteranope, deuteranope, red anomalous with and without darkened red, and protanope (8). The Ishihara Test is cleverly constructed to operate at the juncture of the major defectives with the normal group. It fails at colour blind and anomalous subjects, and in addition a number of normal and colour-weak subjects if strictly applied.

III. COMPARISON WITH THE 'NINE-COLOUR' TEST

Apart from the anomalous and red-green blind subjects, who will be dealt with below forty-five men and fifty-one women who took part in the Nine-Colour Test also did the Ishihara Test, under the carefully standardized conditions mentioned in the previous section. Table 3 shows the numbers of subjects in the various classes in the Nine-Colour

Test and the corresponding numbers of errors they made in the Ishihara Test, giving the range of numbers of errors and the modal numbers, except where there were only two or three subjects in a group. It will be seen that the Ishihara Test gives a very slender indication of the degree of red-green defect among these subjects. The one 'extreme deviant' man who made 21 errors on the Ishihara Test would have been classed as totally green blind by that test alone, and it is a possibility that he was really a green anomalous subject with a sex-linked major defect, whose green deviation was unusually small.

With the exception of this subject it appears that the modal number of errors tends to increase from 1 to 5 with increasing red-green defect, colour weakness having more influence than deviation in producing errors. Since the range of errors can be as great, or almost as great, with subjects who show no colour defects, as with the most extreme deviants (excluding the subject mentioned above and the green anomalous, who will be dealt with later), and the red-green weak, the Ishihara Test could not be used to discriminate between these and the subjects who are absolutely normal. The matter may be confirmed by calculating the mean number of errors made by the absolutely normal subjects (4.25) and by all the other subjects taken together (5.14). The difference between these means divided by its own standard error is $0.89/0.502 = 0.177$, which is not statistically significant. This confirms the result of the Rotating Disks Test, and it also confirms common experience, because nobody has been able to use the Ishihara Test successfully for this purpose, and it has never distinguished between deviant and colour-weak subjects—indeed, if it had been our only guide we should never have suspected such a difference. In addition, the conclusion may be drawn that a person may make up to 10 errors on the

Table 3. *Normal and minor or non-sex-linked defectives in the Nine-Colour Test: Ishihara results*

Type of subject	Nos. of subjects and Ishihara Test errors			
	Men		Women	
	No. of subjects	No. of errors	No. of subjects	No. of errors
Absolutely normal	32	0-8 Mode = 1	19	0-9 Mode = 1
Non-sex-linked defectives:				
Deviant	6	1-8 Mode = 3.5	11	2-11 Mode = 2
Extreme deviant	2	3, 21	3	1, 3, 3
Red-green weak	5	1-7 Mode = 4	18	0-11 Mode = 5

Table 4. *Major defectives in the Nine-Colour Test: Ishihara results*

Type and no. of subjects		Ishihara Test result			
		No. of errors	Darkening of red		
			Dark	Not dark	Indecisive
Protanopes:					
Moderate	2	20, 24	2		0
Extreme	2	24	2		0
Deutanopes:					
Moderate	6	20-24		4	2
Extreme	3	24		3	0
Green anomalous	6	21-24		2	4
Red anomalous:					
Red not dark	1	24		0	1
Red dark	2	22, 24	1		1

Ishihara Test without there being any certainty that suspicions of red-green major defects are confirmed.

Twenty-two of the major red-green defectives who did the Nine-Colour Test also did the Ishihara Test, and the results of this comparison are shown in Table 4. It will be seen that the number of errors made on the Ishihara Test does not distinguish either between extreme and moderate protanopes, extreme and moderate deuteranopes, or between any of these defectives and the red or green anomalous. All made 20 errors or more. In other words, where clear distinctions are found with the anomaloscope, no clear distinctions can be made with the Ishihara Test.

Similarly, the distinctions between protanopes and deuteranopes, between red and green anomalous and between red anomalous with and without darkened red, are also often indecisive on the Ishihara Test. Taking all twenty-two subjects, in six of them the red was darkened, but the Ishihara Test was indecisive for one of these and for 7/16 of those in whom the red was not darkened. It was considered 'decisive' when three of the four places supposed to distinguish protanopes from deuteranopes indicated correctly whether the red was darkened or normal. It has been pointed out elsewhere⁽⁸⁾ that no protanope has been found by the writer in whom the red is as little darkened as the darkest red found in a deuteranope.

The general conclusions of this comparison of the Nine-Colour and Ishihara Tests fully accord with expectations based on the principles on which pseudo-isochromatic tests are constructed. None of the colour differences between the individual spots in the Ishihara Test is subliminal for normal colour vision, and it is doubtful whether a test in which subliminal differences were to be used could be made by colour printing. Every expert in psycho-physical measurements knows, however, as a first principle, that subliminal differences must be available for an efficient test of sensory acuity to be made. This shortcoming of the Ishihara Test is met by arranging the spots of colour in confusing patterns so that some people can read certain figures, others can see the differences between the dots without being able to read the figures, and others again cannot even see the difference between the dots. For subjects with normal, deviant or slightly weak colour vision the differences between 2 and 8 or 10 errors are due to chance factors which are irrelevant to the sensory acuity being tested. For the red-green blind and anomalous subjects however, many or all of the differences between the critical dots may be subliminal, with the result that these subjects make more than 80% of errors in most cases, and can be picked out with some confidence.

IV. COMPARISON WITH THE 'FOUR-COLOUR' TEST

In order to compare the results of the Ishihara Test with those of the Four-Colour Test seventy-six major red-green defectives and forty-three normal subjects were given both of these tests, and their results are summarized in Table 5. The protanopes and deuteranopes are divided into extreme and moderate, as before, and the normal subjects into absolutely normal, deviant and colour weak, according to the principles explained elsewhere⁽⁸⁾, and the Ishihara Test was carried out with the precautions described in section II.

It will be seen that the Ishihara Test fails all extreme deuteranopes and protanopes with 24 errors out of 24. Moderate protanopes and deuteranopes tend to make rather fewer errors, but cannot be distinguished satisfactorily on the Ishihara Test from the subject with extreme defects, because a considerable number of them also made 24 errors, and the

one who made only 4 errors was very exceptional. Apart from him, the smallest number of errors made by a colour blind subject was 13. The anomalous subjects, again, are distinguished neither from extreme nor from moderate protanopes or deuteranopes. Five out of nine red anomalous subjects made 24 errors, and only five out of twenty-five green anomalous subjects made fewer than 17 errors. Hence the Ishihara Test does not clearly separate these groups of major defectives.

The colour weak made an average of 4.5 errors, while the deviants made an average of 3.5 errors with one exception at 8 errors. These types, clearly distinguished by the Four-Colour Test, are not satisfactorily separated by the Ishihara Test, either from each other or from the absolutely normal subjects, who made an average of 3.5 errors (also with one exceptional subject with 8 errors).

Table 5. *Normal and defective subjects in the Four-Colour Test: Ishihara results*

Type and no. of subjects		Ishihara Test result	
		Average no. of errors	Scatter
Protanopes:			
Moderate	3	20.7	18-23
Extreme	15	24	0
Deuteranopes:			
Moderate	19	21	4-24
Extreme	5	24	0
Green anomalous	25	18	9-24
Red anomalous	9	22	14-24
Deviants	18	3.5	0-8
Colour weak	14	4.5	0-12
Normal	11	3.5	0-8

Table 6. *Major defectives in the Four-Colour Test: indecisiveness of Ishihara indication of darkened red*

Type and no. of subjects		Darkening of red		
		Dark	Not dark	Indecisive
Protanopes:				
Moderate	3	0		3
Extreme	15	4		11
Deuteranopes:				
Moderate	19		10	9
Extreme	5		3	2
Green anomalous	25		11	14
Red anomalous:				
Red not dark	6		3	3
Red dark	3	1		2

In general, the conclusion may be drawn again, as in the previous sections, that the Ishihara Test is an effective but indiscriminating cut-out mechanism, which fails all major red-green defectives (with a very few exceptions) and passes all normal, colour weak and deviant subjects (again with very few exceptions). If a failure level of at least 12/24 errors is required, then it may be assumed that few will be failed who ought to pass and extremely few will be passed who ought to fail. It is still an important and difficult question how many colour weak and deviant subjects, who would pass the Ishihara Test, ought to be considered as dangerous or slightly dangerous in occupations involving red-amber-green signal systems, and what sorts of occupations they might have a difficulty in performing in industry. Such problems can be settled only by careful experimental research.

Seventy-six major red-green defectives did the Ishihara Test in this group, and Table 6 shows that for only thirty-two of them did this test decisively distinguish those with darkened red. That is to say, no more than 13/24 deuteranopes, 4/18 protanopes, 11/25 green and 4/9 red anomalous subjects were decisively and correctly placed as having or not having darkened red. It would seem therefore, again, that estimates of the relative numbers of protanopes and deuteranopes in the population, if based on the Ishihara Test, must be very unreliable. This is to be expected, as explained above, because the Ishihara Test does not provide a variable darkness of grey with which to compare the red being tested. In addition, about one-third of the red anomalous subjects have the red end of the spectrum darkened, but the remaining two-thirds of them would not be distinguished by the Ishihara Test in any way from deuteranopes, though their colour vision is completely different.

Finally, two typical green anomalous subjects were tested, both of whom were also extreme blue deviants. The first of these had been tested several times in the R.A.F., and was classed as 'colour blind' on the pseudo-isochromatic tables and as 'defective-safe' on the lantern test. At first he had been passed. The other was passed by the R.A.F. in 1941, and became a pilot. In the laboratory the first made 20 errors on the Ishihara Test, but the second only 9 errors, and, had not some hesitation been noticed in his responses, he would probably have been passed in a quick survey of a class of ninety students. It was easy to show that he was a typically green anomalous subject in the Four-Colour Test. His capacity to read the figures in the Ishihara Test seemed to be due to his blue deviation. When he re-read this test with a pale blue $\frac{1}{2}$ W.-to-daylight filter, which restored the balance in his yellow-blue vision by subtracting yellow, he made 24 errors. In the first of these subjects, however, this explanation did not apply, because he already made 20 errors, without the daylight-blue filter, and so we see the inadequacy of the Ishihara Test in a critical situation. It passed one green anomalous man because he had an extreme blue deviation, which made the pinkish orange dots more visible to him than is usual with anomalous subjects, but with another, almost identical in colour vision, it had a different effect for unknown reasons. If the critic suggests that the fault lay with the Four-Colour rather than with the Ishihara Test, then it must be said that the green anomalous condition is extremely marked and could not possibly be mistaken for any other defect except perhaps that of an exceptionally deviant deuteranope who had a rather small matching range, and such a subject would be even more defective. Undoubtedly the confusion was due to the Ishihara Test.

V. CONCLUSIONS

We may now try to answer the questions asked at the beginning of this paper.

(1) The Ishihara Test can be used successfully to distinguish major red-green defectives from normal, deviant and colour-weak subjects. Major defectives very rarely give as few as 12/24 errors on the test, and most often 18/24 or 24/24 errors. Normal, deviant and colour-weak subjects, however, seldom give more than about 5 errors, and very seldom as many as 8 or more. This has been shown clearly by Murray (6), for the shortened Ishihara Test.

Consequently, a failure level of 12/24 errors will select all the major red-green defectives, with very few exceptions indeed; in this research only two in more than 150 major defectives might have been wrongly passed. They were detected by observing that

they hesitated in a characteristic way, and they were easily shown to be seriously defective by the anomaloscope. This level of failure will, however, include with the major defectives a small proportion of normal, deviant and colour-weak subjects, who are not serious defectives and do not have sex-linked defects. It is more important that the smallest possible number of major defectives should be passed, than that no minor defectives and absolutely normal subjects should be failed, while life and safety depend on good red-green colour vision.

(2) There is certainly a tendency for moderate protanopes and deuteranopes to give a few more correct readings in the Ishihara Test than those major defectives with the most extreme defects. The anomalous often give a few more correct readings than the most extreme red-green blind. This test, however, does not serve to distinguish the moderately from the extremely red-green blind, nor to distinguish either of these groups clearly from the anomalous. In fact, if this were the only test we had to guide us, no clear distinctions of types of major defectives would be possible, and we should have the impression that they were easily confused (which, indeed, is the view taken by many workers). The Ishihara Test fails to distinguish clearly between protanopes and deuteranopes, between the red anomalous and green anomalous, and between the red anomalous with and without darkened red. Nevertheless, Chapanis considers it the best of three similar tests for this purpose (*a*).

(3) No relation could be found between the numbers of errors made by deviant and colour-weak subjects and the degrees of their weaknesses, and absolutely normal subjects could not be distinguished from them consistently by the Ishihara Test.

The most general conclusion is that the Ishihara Test is a cut-out mechanism which operates at the positions of the gaps or junctures between the distributions of major red-green defectives and other subjects. To those who wonder how such a test can be made to succeed, it must be replied that it works only because of the existence of these gaps or junctures. If short peas were generally about 2 ft. high and tall peas about 4 ft., then we could use a piece of elastic 3 ft. long to distinguish between them, provided their variations were distributed in such a way as to overlap as little as those of red-green major defectives and other subjects. It would not make much practical difference if the elastic was stretched a little in making some of the discriminations, but it would be as inadequate a guide to the actual heights of individual pea plants as the Ishihara Test is of the degree of individual colour-vision sensitivities, and it would not work successfully unless there were a gap, juncture or discontinuity between the heights of short and tall peas.

VI. SUMMARY

In order to study the validity of the Ishihara Test for colour blindness, the results for normal and defective subjects on this test were compared with their results on three tests based on the Rayleigh Equation.

Twenty major red-green defectives and sixty-seven normal and colour-weak subjects did the Ishihara Test and a test of red and green using Hering's coloured papers mounted on Marbe's rotators. Twenty-two major red-green defectives and ninety-six normal, deviant and colour-weak subjects did the Ishihara Test, and a red-green anomaloscope test in which monochromatic colour filters were used; and seventy-six major red-green defectives and forty-three normal, deviant and colour-weak subjects did the Ishihara Test and an improved form of this anomaloscope test. In all 206 normal, deviant and

colour-weak subjects and 118 major red-green defectives were studied. The techniques of the anomaloscope tests, and the characteristics and frequencies of the different types of major and minor defectives have been dealt with in another paper (8).

The general conclusions were: (1) The Ishihara Test gave a satisfactory cut-out of major red-green defectives on a failure level of 12/24 plates. If the level were lowered a few major defectives would be missed, and this would be serious while life and safety depend on good red-green vision. (2) The Ishihara Test failed to make any clear distinctions between the extreme and moderate colour-blind subjects, between protanopes and deutanopes, between red and green anomalous and between these and the colour blind. It is an almost indiscriminating cut-out for major defectives. (3) The Ishihara Test could not be used to distinguish between any kinds and degrees of minor defects revealed by the other tests.

REFERENCES

- (a) CHAPANIS, A. (1949). Diagnosing types of colour deficiency by means of pseudo-isochromatic tests. *J. Ophth. Soc. Amer.* XXXIX, 242-9.
- (1) COLLINS, M. (1929). The Rayleigh equation with rotating disks. *Brit. J. Psychol.* XIX, 387-93.
- (2) COLLINS, M. (1937). Tests in common use for the diagnosis of colour defect. *Brit. Assoc. Ann. Rep.* pp. 208-26.
- (2a) GATES, R. R. (1946). *Human Genetics*, Ch. vi. London: Macmillan.
- (3) GEDDES, W. R. (1946). The colour sense of Fijian natives. *Brit. J. Psychol.* XXXVII, 30-6.
- (4) GRIEVE, J. (1946). Incidence of defective colour vision. *Nature, Lond.*, CLVII, 376.
- (5) MEDICAL RESEARCH COUNCIL (1933). Reports of the committee on the physiology of vision. *Spec. Rep. Ser. Med. Res. Coun., Lond.*, no. 185, 58 pp.
- (6) MURRAY, ELSIE (1948). Mass-testing of colour vision; a simplified and accelerated technique. *Amer. J. Psychol.* LXI, 370-85.
- (7) PICKFORD, R. W. (1944). The Ishihara Test for colour blindness. *Nature, Lond.*, CLIII, 656.
- (8) PICKFORD, R. W. (1949). Individual differences in colour vision and their measurement. *J. Psychol.*, XXVII, 153-202.
- (9) RIDDELL, W. J. B. (1949). Colour Vision and Industry. *Proceedings of Royal Philosophical Society of Glasgow*, LXXIII, 83.
- (10) VERNON, P. E. & STRAKER, A. (1944). Distribution of colour-blind men in Great Britain. *Nature, Lond.*, CVII, 690.

(Manuscript received 20 May 1949)

THE OBJECTIVE MEASUREMENT OF ATTITUDES

By R. B. CATTELL, E. F. MAXWELL, B. H. LIGHT AND M. P. UNGER

From the University of Illinois

I. THE DEFINITION OF AN ATTITUDE

The present article is one of a series of five concerned with theoretical and practical advances in the objective measurements of interests and attitudes^(2,4,7,8). There is considerable evidence that the measurement of attitudes by self-evaluative and self-declaratory methods is unreliable and invalid, not only in the prediction of more substantial behaviour, but even in the prediction of verbal behaviour upon some other occasion or in some other context^(12,13). Even if this evidence did not exist, however, psychological analysis would have sufficed to indicate the desirability of shifting the assessment of attitudes from gross verbal methods dependent upon the honesty and insight of the subject to methods of a more indirect kind based upon our growing knowledge of psychodynamics.

Before describing the present experiments, it is necessary to define more precisely what is meant by an attitude. The attitude is a prototype of all dynamic traits, in that it involves an intended direction of action with respect to an object and is aimed ultimately at satisfying certain basic drives. Any attitude needs to be defined initially by five aspects, which can be summarized in the paradigm: '(I) In these circumstances; (II) I; (III) want so much; (IV) to do this; (V) with that.' In this statement: (I) defines the stimulus situation with reference to which the attitude is stimulated and evoked as well as maintained; (II) defines the organism bearing the attitude; (III) the strength of interest in the course of action indicated; (IV) the kind of action indicated; and (V) the object with which the attitude is concerned. Sometimes the stimulating situation and the object with respect to which action is to be taken are the same.

The above analysis has been developed in the ergic theory^(3,4,7) of attitude measurement according to which an attitude can be expressed for purposes of analysis and calculation as a vector quantity. In this quantity the length of the vector represents the strength of the desire, that is to say, the interest in the defined course of action, while its direction represents the *dynamic composition* of the satisfactions to be gained by that course of action. The ergic co-ordinates by which attitude vectors are placed represent the various ultimate drive satisfactions which may be obtained through any attitude. Since little work has yet been done on the number and nature of human drives (ergs), the present theory assumes that appropriate factor analytic procedures will yield as factors those primary drives which obtain satisfaction in common attitudes. This assumption is tested in one research in the present series⁽⁷⁾. Thus the projection of a given attitude upon the various co-ordinates will define the extent to which certain ergs, e.g. hunger, sex, self-assertion, pugnacity, commonly gain expression through that particular attitude.

The ergic theory⁽⁴⁾ implies that the measurement of attitudes cannot be adequately covered by the existing convention^(12,20) of assigning a score on a pro and con scale with respect to an object. For an attitude to an object is far richer than a single dimension can express, and is better defined by all those basic drive satisfactions to which this particular

attitude to the object leads. One can, of course, correctly speak of a pro-con scale with respect to a defined course of action. But a person may utilize the same object for many different courses of action so that, for this reason, as well as on account of the dynamic complexity, mentioned above, of a single course of action, it is meaningless to speak of a pro-con attitude towards an object.

The ergic theory leads to a formula for the strength of the attitude akin to that used in the specification equation (3) for expressing a particular skill in terms of primary abilities, as follows:

$$I_{ij} = S_{1j}E_{1i} + S_{2j}E_{2i} + \dots + S_{nj}E_{ni} + S_jE_{ji},$$

where I is the strength of interest of the individual i in the course of action defined by the attitude j . The S 's are the factor loadings, which in this case we shall call the dynamic *situational indices* (3,7), defining the extent to which the various ergs or drives, E_1, E_2 , etc., are involved (for the average member of the population) in determining the course of, and gaining satisfaction through, the action concerned.

Actually, it is not necessary to subscribe to the theory concerning this particular manner of measurement of direction, or, indeed, to know the co-ordinates involved in fixing the direction of the attitude, in order to utilize the methods to be discussed here. It suffices only that the direction of the attitude be *rigidly fixed in terms of behaviour* with respect to the object. Within this framework of properly defined and fixed directions of behaviour, we may then set out to measure the length of the vectors, that is to say, the strength of the interests in the given courses of action. The problem of determining the direction of the attitude vectors, that is to say, the measurement of the S 's in the above equation, is discussed fully elsewhere (7). The present article, therefore, is concerned solely with determining this strength of interest in the course of action which is verbally or otherwise defined beforehand.

II. OLD AND NEW METHODS EMPLOYED IN THIS EXPERIMENT

The experiment aimed at taking a wide and representative set of attitudes of a kind likely to be of practical interest and utility, not only to social psychologists but also to clinicians and others. It planned to measure these attitudes by a variety of methods including, on the one hand, some conventional 'tried' methods and on the other, a variety of entirely new methods founded upon the application of psychodynamic principles.

Some twenty-five methods were worked out for the total research on objective measurement of dynamic traits, but they are only listed in bald fashion below (Table 1), having been described fully enough for experimental purposes elsewhere (8). Twelve of them have been examined for validity in a report (8) parallel to the present one: ten, partly overlapping with these twelve, are reported upon fully here; a few methods remain to be reported upon.

Only the ten methods (starred items) that were used in the present experiment will now be described, with respect to experimental details and rationale, as follows:

(1) *Measurement of Expenditure of Time*. The individual was asked to keep an exact record of his expenditure of time upon the given activity in each of 2 weeks, a month apart and disturbed by no holiday season. A reliability coefficient was obtained by a comparison of the 2 weeks. This method has received a trial in Thorndike's studies of interests (17).

(2) *Measurement of Money Expended*. The individual was asked similarly to take 2 weeks, a month apart and clear of any special holiday season, and to record his ex-

Table 1. *List of experimental approaches designed for the attitude and dynamic trait measurement research*

(1) Expenditure of Money on an Activity*	(15) Measurement of Phantasy
(2) Expenditure of Time on an Activity*	(16) Measurement of Projection*
(3) Opinionnaire Methods ('Classical', verbal, questionnaire method)*	(17) Measurement of Ego Defence Dynamisms Generally
(4) Preference Methods (Choice Box Situation)*	(18) Measurement of Fluency on a Particular Topic
(5) Measurement of Spontaneous Attention	(19) Measurement of Speed of Reading
(6) Measurement of Immediate Memory*	(20) Measurement of Work or Endurance (Obstruction Method)
(7) Measurement of Reminiscence	(21) Measurement of Psychogalvanic Response*
(8) Measurement of Distraction Effects	(22) Measurement of Pulse Rate and Autonomic Changes*
(9) Measurement of Retro-active Inhibition Effects	(23) Measurement of Metabolic Rate
(10) Measurement of Action Information*	(24) Measurement of Muscle Tension
(11) Measurement of Speed of Decision	(25) Measurement of Writing Pressure*
(12) Measurement of Level of Skills	
(13) Measurement of Misperception (Illusion)	
(14) Measurement of False Belief (Delusion)	

penditure on the particular interest activity concerned for the whole of each week, the reliability coefficient being calculated by correlation of the 2-week periods. This method seems to have been employed systematically only by Thorndike (18).

(3) *Measurement of Preference.* The individual was given a list of fifty attitude activities, the ten attitudes with which the present experiment is concerned being embedded among them. He was then asked to make every possible comparison of these activities in the matrix of cells (1225 in number) presented by the intersections between the row of fifty attitudes and the column of fifty attitudes. In each paired comparison he stated which of the two activities he would prefer to do, thus reacting analogously to the animal in the classical 'choice box' (20) motivation test. The score for any one attitude was therefore the number, out of forty-nine comparisons, in which the given attitude was the preferred activity. A reliability coefficient was obtained by the split-half method among the fifty items. The fifty attitudes involved, only twelve of which are reported upon here, have been set out elsewhere (7).

(4) *Measurement of Immediate Memory.* A series of brief statements were presented tachistoscopically, ten to each of the attitudes. For each attitude five of the statements concerned conditions or objects which would facilitate the expression of the attitude, while five of them referred to activities or objects which might frustrate the expression of the attitude. This subdivision into facilitating and frustrating references was made in order to test a further hypothesis that ideas connected with the attitude in a frustrating fashion might have different effects from those connected with the attitude in a facilitating manner. As examples of 'facilitating' and 'frustrating' presentations we may take two from those used for attitude 6 (wanting success in one's career), viz. 'Success in career assures happiness' and 'The successful career man is always selfish'. However, the statements for all fifty attitudes were intermixed in presentation and the main purpose was to see whether ideas, of any kind, having connexion with an attitude which the person strongly held would be preferentially remembered in comparison with those connected with an attitude which interested him less. After the exposure of twelve statements at 6 sec. intervals and a pause of 25 sec. the individual was asked to recall in 30 sec. all that he could remember of 'the phrases, statements or ideas presented in this last period'. This was repeated forty-two times, a new disk of twelve phrases being set up each time. This session, and these phrases, were also used at the same time for the P.G.R. measurements described below. Credit for the recall was given when the essential idea

of the item was reiterated regardless of verbal form. A score was thus eventually obtained of the accumulated items recalled with respect to each of the attitudes.

(5) *Measurement of Action Information.* This took the form of an ordinary multiple choice test. Ten information items were presented for each attitude, the information required being not concerned with the *object* (except accidentally), but with the implementation of the course of action connected with the object, i.e. the behaviour necessary to satisfy the attitude. The subject was asked to leave no item unanswered but to guess when in doubt. This test was found to be far more difficult to construct for some attitudes than others. For example, the knowledge required to implement the attitude of ambition towards a career in engineering is much more definite than that required to implement the desire for greater indulgence in alcohol or to show more affection towards one's child. A typical example is:

'In the $\left\{ \begin{array}{l} \text{Orange} \\ \text{Sugar} \\ \text{Cotton} \end{array} \right\}$ Bowl game of January 1st, 1948 $\left\{ \begin{array}{l} \text{Georgia} \\ \text{Michigan} \\ \text{S.M.U.} \end{array} \right\}$ defeated $\left\{ \begin{array}{l} \text{Alabama} \\ \text{California} \end{array} \right\}$,

used for attitude 22; "wanting to follow football games as a spectator".

(6) *Measurement of the Psychogalvanic Reflex Response.* This was measured as percentage change of resistance to the tachistoscopic exposure of the same items as were used in the immediate memory test. Again it was possible, apart from obtaining a score for all the items for a given attitude, to obtain separate scores for the deflexions to those ideas which facilitated the expression of the attitude and those items which frustrated it. The measurement of the psychogalvanic response was taken with the usual technical precautions as described elsewhere(1).

(7) *Measurement by the Opinionaire Method.* This was the classical opinionaire method as described by Thurstone(19) and others. A number of statements were obtained, and ten were eventually chosen for each attitude, of such a nature that they gave a graded series of conviction of statement between an opinion entirely opposed to, or indifferent to, the attitude on the one hand, and an opinion expressing strong interest in the consummation of the purpose of the attitude on the other. No elaborate scaling methods were attempted in the present study. The items were simply ranked in order according to the extent of their favouring satisfaction of the attitude, and the individual was scored on the opinions that he was willing to endorse.

(8) *Measurement of Writing Pressure.* The theory behind this test is that a person in replying to a question concerning his interests is likely to exert a stronger pressure in writing yes or no for those attitudes where he has stronger emotional involvement. Three positive and three negative statements were made up with respect to each attitude, and the individual was asked whether he agreed or disagreed. The strength of the pressure was registered pneumatically and transmitted to a tambour. The scoring in this test was ipsative (see below). That is to say, since the *mean* handwriting pressure of one individual differed substantially from that of another, his scores for the individual statements were scored in terms of the average S.D. of his own total population of pressure scores for all the individual statements. Since only eighteen statements were involved in all, the scoring was necessarily coarse. Unfortunately, the decision was made in this experiment to add pressures (for each individual on each attitude) regardless of the affirmativeness or negativeness of the response, which may account for some of its failure.

(9) *Measurement of Pulse Pressure.* This test was based on the theory that the subject would react with some slight increase of systolic blood pressure when he was presented

tachistoscopically with a statement which provoked a stronger conative reaction on his part. Five 'facilitating' and five 'frustrating' observations were offered at 5 sec. intervals, tachistoscopically, for each attitude, *S* being asked for no remark or reaction to them. The cuff pressure of 80 mm. of mercury on the brachial artery was released after each series of ten statements. The ipsative scoring (see below and (3)) in this particular measure was based on the population of only thirty items (for three attitudes), which necessarily means that this test is more coarsely scored than some others, in which the attitude strength is expressed relative to the S.D. of fifty attitudes. The height of the systolic surge was measured on the record card by vernier callipers for the heart beats during the 5 sec. following each exposure and the average for these beats was obtained. The apparatus used was the Cameron Heartometer developed by Cureton and others (11).

(10) *Projection-Phantasy*. This was a projection test only in the loose sense in which 'projective' (6) is frequently used. Actually it is a 'choice of phantasy' test. A hundred unfinished sentences were presented (ten for each attitude) in groups of ten (one in any ten representing each attitude). Each began by indicating a certain type of attitude satisfaction, e.g. 'Mr *X* often went to Art Galleries because...' and *S* was asked to choose which five he would like to complete (after reading all ten of them). It was not his completions but the nature of the themes he chose to complete that was recorded as a measure of his attitude interests.

III. DESIGN OF THE EXPERIMENT

The above measures were applied to a population of 200 young men, mainly students, and all married, since it was desired to include powerful attitudes having to do with the family. The correlations for any one attitude, however, are on a population of forty, since it became necessary, owing to demands on the subject's time, to divide the population into five groups, each of which shared certain methods with others while having some methods restricted to itself.

A method of measuring attitudes can be said to be good only if it is valid and applicable for attitudes of diverse kinds. It was an essential part of our design to set up a very representative set of dynamic traits, including attitudes to the family, to the self, to career and hobbies and to social, religious and political questions. Measures were taken on fifty attitudes in all, scattered over the fifteen main categories in Cattell's interest test (5). The full fifty had to be included in order to obtain an adequate basis for the ipsative measurements, but the data were fully worked out in the present experiment only for the following twelve attitudes. These have the identifying numbers used in the full set of fifty attitudes set out elsewhere (7).

Table 2. *List of attitudes used in the experiment*

- (1) I want to play more indoor sociable games, such as card games.
- (2) I want to spend somewhat more on drinking and smoking than I am now able to do.
- (6) I want to become proficient—if possible to excel my colleagues—in my chosen career.
- (10) I want more time to enjoy sleep and rest.
- (11) I want to listen to music.
- (16) I want to know more science.
- (19) I want to see organized religion maintain or increase its influence.
- (22) I want to attend football games and follow the fate of teams.
- (30) I like to see a good movie or play every week or so.
- (34) I want to get my wife the clothes she likes and to save her from the more toilsome household drudgeries.
- (36) I want to be smartly dressed, with a personal appearance that commands admiration.
- (44) I want to feel that I am in touch with God, or some principle in the universe that gives meaning and help to me in my struggles.

It was planned to validate the measurement methods by 'internal validation' (3), the criterion being the consensus, i.e. the pooled score, on the best methods. The set of best methods can to some extent be chosen, *a priori*, in so far as some methods, to wit, expenditure of money and expenditure of time, are by their very nature to be taken as criteria. Others have already been shown experimentally to have some degree of validity, notably, immediate memory (2), information (2,16) and the psychogalvanic reflex (1,9,10,15). The design called for these 'tried' methods to be applied to all attitudes, to provide a solid core for validation, while each of the more 'explanatory' methods—(7)–(10) above—was applied to only one or two attitudes, as indicated in Table 5, below. However, the membership of this core, for the purpose of ultimate validation of each method, was left to the arbitrament of experiment. In the end the validity of a method was its mean correlation with all other methods that proved to have appreciable correlations with their fellows, except that, by the necessary pre-designing of the experiment only one 'exploratory' method, however good the other exploratory methods, could appear in the validating pool at one time.

In sum, the performances of the subject consisted of keeping a 2-week record of time expenditures; a 2-week record of money expenditures; completing, in about 2 hr., a set of 1225 paired preference comparisons; completing a 500-item information test; sitting for two 2-hr. sessions connected into the P.G.R. viewing a tachistoscopic exposure of 500 verbal stimuli, being asked at intervals to recall these, and, finally, having a special session with one of the 'exploratory' methods.

Discussion of scoring involves the recognition that through all dynamic trait, attitude and interest measurement there runs the dilemma of *ipsative* (3) versus *normative* standards, as yet only briefly mentioned here. Is an interest manifestation to be scored relative to the individual's total interest-energy—which may differ considerably from person to person—or relative to the mean and scatter of interest manifestations among individuals constituting the population? Wherever the manifestations are actual behaviour, truly comparable from person to person, the latter method, which is in line with that followed in measuring abilities, is to be preferred. But wherever the expression of interest is mediated by some second factor, at the peculiar level of endowment of the particular individual, it seems desirable to score ipsatively and to express the given interest *in terms of its rank order (or standard score) among all the given individual's interests*.

For example, the average psychogalvanic reactivity among individuals differs tremendously, and we cannot assume that these differences express differences in 'total' interest. They are more likely to express some second factor of general emotionality or some physiological reactivity. Consequently, when the individuals' P.G.R. scores are obtained for a given attitude they must be expressed for each person relative to his own general P.G.R. reactivity, i.e. ipsatively, before inter-personal comparisons with respect to that particular attitude can be made.

The scores that were put in ipsative form before correlating were those of the P.G.R., the Immediate Memory test, Writing Pressure and Pulse Pressure. Preferences, Projection and Expenditure of Time were naturally in this form, since the total preference and projection scores were the same for all and 'there are only 24 hr. in a day' for everyone. Some claim could be made for assessing money expenditure relative to the individual's total financial resources, and information relative to his general intelligence, but these and the opinionaire were left in 'interactive' (3) units and treated normatively. The fact that

the reliabilities and validities of the Time and Money expenditure methods are so similar and their intercorrelation good shows that absence of ipsative rescaling, in these debatable borderline circumstances, at least, produces no radical difference in results.

IV. RESULTS

The reliabilities of the methods need first to be examined. Those for the core of 'tried' methods rest each on some eight to twelve different attitudes, as set out elsewhere (8), each with forty cases in the correlation, so that the standard errors of the correlations given below range from 0.005 to 0.075.

Table 3. *Reliabilities of 'Tried' methods*

Expenditure of Time	0.94	Psychogalvanic Reflex	0.85
Expenditure of Money	0.94	Information	0.59
Preference Method	0.90	Immediate Memory	0.50

These are split-half reliabilities, based generally on five items in each half, and corrected to the full length of ten items or the total observation period. Although the peculiar difficulties of making up certain types of tests for certain attitudes, e.g. the obscurity of the special information likely to be possessed by a person with unusually strong interest in sleep, caused a few wide divergencies, the reliabilities for methods, in general, represent pretty consistent central tendency.

The reliabilities for the methods tried out, each on only one or two attitudes, were: Writing Pressure (attitude no. 19), 0.77; Opinionaire (attitude no. 22), 0.37; (attitude no. 30), 0.37; Projection (attitude no. 6), 0.60; (attitude no. 10), 0.13; and Pulse Pressure (attitude no. 11), 0.97. (Attitude numbers as in Table 2, above.)

In general, the reliability of these tests is good enough to permit the experiment to yield fair estimate of their validities, but Information, Immediate Memory, Opinionaire and Projection certainly leave scope for improvement of the consistency of items, and these methods must have their validities considered in the light of these only moderate reliabilities.

The validities of the 'core' methods, as indicated above, are estimated as their mean correlations with the pool. The correlations in the pool are set out in Table 4 below. Each entry is the mean of correlations on from eight to twelve attitudes and thus represents population of from 320 to 480 cases, while the validity coefficients at the bottom, being the means of five such coefficients, may be considered as based on from 1600 to 2400 cases. The standard errors are estimated from the lower figure, because of the approximations involved in the above reasoning.¹

The validities of the remaining methods are shown with more specificity (since it is relevant to take account of the particular attitudes on which they were tried), in Table 5.

It will be observed that two methods, Immediate Memory and the P.G.R., have been

¹ Initially the r 's were averaged through Fisher's Z -function, which proceeds on a population of $N-3$ cases, e. 37 in this instance. The equivalent of a correlation derived from correlations on eight independent, chance-sampled groups of 40 is therefore one on a population of 296. But each validity r (bottom row, Table 4) is the mean of five such r 's, though these, it is true, represent the results of five different experimental assessments of correlation (the correlation with a criterion) on the same population. The above standard errors, based on 300 cases, therefore represent an approximation, using a figure close to the lower limit of the possible population involved in each.

dropped from the six 'core' methods, on account of having the lowest validities in Table 1 above. Elsewhere (8) it is shown that the relation of Immediate Memory and of the P.G.R. to interest is probably more complex than a simple linear relation, and that a mixture of 'facilitating' and 'frustrating' verbal stimuli (with respect to the attitude being measured) is probably of less diagnostic value, in these methods, than a test composed of items of one kind only. At least, for the present, it seems best, therefore, to assess the exploratory methods against a more conservative core of methods, omitting these two.

Incidentally, the opinioaire approach, which is the classical verbal method of attitude assessment, may seem wrongly labelled under 'exploratory' methods—as indeed it is—but it cannot be placed in the core of 'tried' methods unless it proves itself, by the methods and standards of this research, to have highly significant correlations with the pool of all methods.

The significance of the actual mean validity coefficients of Table 5 is again calculated on the assumption that the standard error of the mean of n correlation coefficients each on

Table 4. *Validities of basic methods*

	1	2	3	4	5	6
(1) Time Expenditure	—	—	—	—	—	—
(2) Money Expenditure	0.47	—	—	—	—	—
(3) Preference method	0.21	0.25	—	—	—	—
(4) Information	0.16	0.15	0.16	—	—	—
(5) Immediate Memory	0.01	0.03	0.13	0.01	—	—
(6) Psychogalvanic Reflex	0.08	0.04	0.03	0.04	0.15	—
Mean correlation with pool	0.19	0.19	0.16	0.10	0.05	0.07
Maximum s.d. of correlation	0.02	0.02	0.02	0.02	0.025	0.025
Mean correlation with pool of four most valid methods	0.28	0.29	0.21	0.16	0.02	0.05

Table 5. *Correlations of exploratory methods with core methods*

	Attitude no.	With preference	With time	With money	With information	Mean	s.d.
Writing Pressure	19	0.28	0.15	0.04	-0.06	0.10	0.08
Projection (1)	6	0.15	0.01	—	0.20	—	—
Projection (2)	10	0.31	0.26	—	0.20	0.21	0.06
Opinionaire (1)	22	0.39	0.16	0.08	-0.01	—	—
Opinionaire (2)	30	-0.01	0.11	-0.09	0.33	0.12	0.06
Pulse Pressure	11	-0.04	0.11	-0.08	0.15	0.04	0.08

a population of N is equal to that of a single correlation on a population of somewhat less than $n \times N$. By this criterion the Projection Method has a validity significant at the 1% level; the Opinionaire Method has a validity significant at the 5% level and the other two methods are no good.

This bald statistical statement must be considered in the light of the following miscellaneous considerations. The Opinionaire tests had an unusually low reliability for such tests. Corrected for attenuation the Opinionaire method would be as significantly valid as any here examined. The scoring of the Writing Pressure test did not take account of whether the pressure was associated with affirmative or negative responses. It might well be valid when scored more analytically. Although the Pulse Pressure method yields nothing at all, our qualitative observations indicate that variation on the method might yet reward further research. In the first place, we observed significant pulse pressure changes in response to specific stimulus statements that *frustrated* attitudes. Correlations of -0.20 and -0.13 were obtained, also, with magnitude of P.G.R. reactions to *facilitating* state-

nents. Possibly the Pulse Pressure Method results should be separately scored for facilitating and frustrating stimuli, the values being added with opposite sign. Experience also suggests that the period of observation after each stimulus should be increased to 10 sec., and that the height of the dicrotic notch, pulse rate and some other features of the curve additional to pulse pressure might prove to be significant.

V. SUMMARY

1. On the assumption that advance in the measurement of motivation, and dynamic traits generally, requires not so much concentration on statistical scaling refinements of present verbal methods as more psychological ingenuity in devising varied methods based on psychodynamic principles, some twenty-five methods, mostly new, have been suggested.

2. The examination of sixteen of these methods, ten being described in the present paper, shows that nine of them—Time Expenditure, Money Expenditure, Preference, Information, Speed of Decision, Distraction, False Belief, Opinionaire Methods and Protection—have significant validity. Two—Psychogalvanic Reflex and Immediate Memory—have promising validity but need re-analysing in regard to curvilinear relationships and the effect of specific ego threat in the stimulus. Five methods—Misperception, Fluency, Speed of Reading, Writing Pressure and Pulse Pressure—failed, as administered here, to show any validity, but suggestions for improving the last two have been made.

3. Most of the correlations found, either with individual methods or with that pool of 'tried' methods which showed most intercorrelation, were quite small. However, our object was to explore widely for tests of *any* validity, and it must be remembered that the scales used would need to be increased in length tenfold to equal the length and reliability of, say, typical intelligence tests. Moreover, in this exploratory study no systematic item analyses were undertaken in setting up the tests. With the experience now gained in constructing tests in this new area the experimenters are in a position to set up items with more psychological skill and such as would have decidedly greater item validity.

4. If the present validities are corrected for attenuation, by allowance for the present imperfect reliabilities, validity coefficients in the region of 0.3–0.5 are obtained. The improvements suggested in (3) above are likely, however, to raise the validity more than the reliability and to produce tests which, if lengthened, should have validities above 0.5. Nevertheless, the present information suggests that a sufficient validity will not be obtainable by any one method and that the objective measurement of attitudes will need a battery employing, say, four to six different methods. In other words, each method is measuring, besides that dynamic strength of attitude measure which it shares with other methods, an appreciable amount of some manifestation of attitude that is peculiar to itself. The attenuation correction shows that this is not merely error of measurement and it becomes of scientific interest, therefore, to discover just what extraneous influences are entering into each method.

For the present, however, concentration on the validity is more important than concentration on the non-validity. Only through exploration of new methods, and the raising, by further research concentrated on detail, of the validity demonstrated in the above methods, can psychology hope to provide those objective attitude measures sorely needed to permit progress in economics, social psychology and the social sciences generally.

REFERENCES

- (1) CATTELL, R. B. (1929). Experiments on the psychical correlate of the Psychogalvanic Reflex. *Brit. J. Psychol.* xix, 357-86.
- (2) CATTELL, R. B. (1935). The measurement of interest. *Character and Personality* iv, 147-69.
- (3) CATTELL, R. B. (1946). *The Description and Measurement of Personality*. New York: World Book Co.
- (4) CATTELL, R. B. (1947). The ergic theory of attitude measurement. *Educ. and Psychol. Meas.* vii, 221-46.
- (5) CATTELL, R. B. (1948). *A Guide to Mental Testing*. London: University of London Press.
- (6) CATTELL, R. B. (1949). Principles of construction of apperceptive or projective tests of personality, Chap. 2 in H. H. Anderson's *Projective Methods*. New York: Wiley.
- (7) CATTELL, R. B. Ergic structure in man, as inferred from the measurement of attitudes. (In the Press.)
- (8) CATTELL, R. B., HEIST, A. B., HEIST, P. A. & STEWART, R. G. The objective measurement of dynamic traits. (In the Press for *Educ. and Psychol. Meas.*)
- (9) CHANT, S. N. F. & SALTER, M. D. (1937). The measurement of attitude toward war by the galvanic skin reflex. *J. Educ. Psychol.* xxviii, 281-9.
- (10) COLMAN, R. D. & McCRAE, C. R. (1927). An attempt to measure the strength of instincts. *For. Educ.* v, 171-81.
- (11) CURETON, T. K. (1947). *Physical Fitness Appraisal and Guidance*. St Louis: Mosby.
- (12) DROBA, D. D. (1932). Methods of measuring attitudes. *Psychol. Bull.* xxix, 309-23.
- (13) McNEMAR, Q. (1946). Opinion-attitude methodology. *Psychol. Bull.* xliii, 289-374.
- (14) PASCAL, G. R. (1943). Handwriting pressure, its measurement and significance. *Character and Personality* xi, 234-54.
- (15) SMITH, W. W. (1922). *The Measurement of Emotion*. London: Kegan Paul.
- (16) SUPER, D. E. & ROPER, E. S. (1941). An objective technique for testing vocational interests. *J. Appl. Psychol.* xxv, 487-98.
- (17) THORNDIKE, E. L. (1937). How we spend our time and what we spend it for. *Sci. Monthly*, xliiv, 464-9.
- (18) THORNDIKE, E. L. (1937). What do we spend our money for? *Sci. Monthly*, xlv, 226-32.
- (19) THURSTONE, L. L. (1929). The theory of attitude measurement. *Psychol. Rev.* xxxvi, 221-41.
- (20) WARDEN, C. J. (1931). *Animal Motivation*. New York: Columbia Univ. Press.

(Manuscript received 1 June 1949)

PUBLICATIONS RECENTLY RECEIVED

Social Adjustments in Old Age. By OTTO POLLAK. Social Science Research Council, Bulletin 59, New York. 1948. Pp. xi+199. \$1.75.

This is admirably called 'A research planning report', for that is exactly what it is. Planning for research must of necessity be extremely general in character. Once basic investigations really begin they are apt, if they are fruitful, to go off in all sorts of un contemplated ways and they must have freedom to do this if they are not going to be merely a way of illustrating what is already predetermined. Dr Pollak has presented an extremely persuasive plea for a fully organized attack upon problems of old age regarded as all centring about some special features of adjustment within or to social communities. Demographers, psychologists (both experimental and social), economists, and sociologists are all going to take part. Any available methods will be used, provided only they are adequately checked up against other available methods; but perhaps with a strong predilection for the survey, both of the more objective, environmental kind and of attitudes and opinions. As a programme it is excellent, within the limits of the notion of 'adjustment' most comprehensive, and in general unexceptional. One wonders, indeed, whether the whole thing might not move ahead a bit quicker and to more genuine purpose if an equal amount of energy and ability were devoted to finding some 'soft' starting-point and really setting to work. As it is somebody will have to make a start somewhere within this plan, for it is altogether too comprehensive and varied for a beginning to be made at every point of it. And then, very likely, something will speedily turn up which will break right outside the scheme. However, if plans of this kind really are needed this is a remarkably well-thought out specimen.

F.C.B.

Les Réactions chez les Nouveau-Nés avant et après le Repas. By LIN CHUAN-TING. Louvain: E. Nauwelaerts. 1949. Pp. 72.

The research reported in this excellent monograph was concerned with the movements that occur in states of satisfaction and of dissatisfaction in very young infants. Slightly modifying an earlier classification of bodily movements, the author first demonstrated that the movements of infants of 1-10 days old could be accurately recorded for short periods by a trained observer without the use of instruments. A high measure of consistency was found in the observations made in successive minutes and on two different days. He then studied the movements of upwards of 500 infants at three periods: just before a feed, midway between two feeds, and just after a feed. Correlations were calculated between the relative frequencies of the various movements in the three periods. It was found that the period of hunger and dissatisfaction showed the same movement pattern, but in an intensified form, as that observed in the neutral period between feeds, but that in the state of contentment after feeds a different pattern of movement appeared. It was characterized by an increase, relative to other bodily movements, of movements of hands, fingers, feet and toes, pushing the tongue out, opening the eyes, moving the eyes, slight twistings of the body, and yawning. The author sees these movements as part of the slackening of tension noted by other investigators. The state of contentment after being bathed produced a similar distribution of movement frequencies. The distress caused by vaccination was accompanied by the same distribution of movements as that found in hunger.

Mr Lin's conclusion is that in the infant's earliest days there occur neither the specific emotional responses indicated by Watson nor the undifferentiated emotional excitement suggested by Bridges, but two distinguishable emotional reactions: satisfaction and dissatisfaction. The care and thoroughness of the research and the significance of its findings for general theory make it noteworthy. It would be lamentable if Mr Lin's further merits of conciseness and unpretentiousness led to the importance of his work being overlooked.

Fijian Village. By BUELL QUAIN. With an Introduction by RUTH BENEDICT. University of Chicago Press and Cambridge University Press. Pp. xvii+459. Illust. 27s. 6d.

The death of Buell Quain in 1939 was clearly a distinct loss to social anthropology, including that branch of it most closely related to psychology. A young American from Columbia University, he spent 10 months in field research in a small inland community in the island of Vanua Levu in the Fiji group. This book,

which gives only part of his results, analyses the structure and organization of the community—under its pseudonym of Nakoroka—on the basis of a detailed observation of the behaviour of almost every one of the hundred inhabitants. Their alinement in households, status groups, land-holding groups, moieties and families is carefully described, and their intricate relations in the various aspects of daily life. The analysis is enlarged by sections of reconstructed life histories. The result is to provide a very vivid and convincing account of a complex structure in which moiety membership and membership of (or exclusion from) the chief class are dominant principles. It would appear from Quain's account that there is no clear-cut lineage system, though his references to the land-owning groups suggest that there is an extended-family structure which he does not bring out. What is demonstrated extremely well is the range and effect of individual choice in forming relations in kinship, land-holding, and ceremonial class ('caste'). The individual chooses in a large measure those with whom he will have the most effective ties of kinship. A man selects from a number of close female kin of the right kind one to be his 'special' sister, to act as his confidant and helper, and to serve, with her children, as the recipient of his special services and gifts. Every individual, it is said, must discover for himself which of his parental ties, with his mother or his father, is the stronger and must so decide his residential and land affiliations. This emphasis on the freedom of the individual against the background of the rules and values of the society leads the author to an elaborate examination of motives for action. Point is made, by Quain and by the late Ruth Benedict in her useful introduction, of the contradiction in ideals offered to the individual by his society. Forbearance, sweet temper, generosity, co-operation win regard—but so also do aggressiveness, acquisitiveness, ruthlessness. Faced by these two incompatible sets of goals, it is argued, the Nakorokan is left without any objective standard of public conduct. His rank and the esteem in which he is held depend on 'subjective judgement and random proof'. The result, it is argued, is anxiety about status, bafflement and poverty of courage. These conclusions do not seem to be fully borne out by the evidence. Most Nakorokans seem to display a fair degree of equanimity in Quain's empirical accounts. And the incompatibility of goals, the split in public judgement, is surely a characteristic of social existence anywhere; one fundamental social dilemma is to decide between aggressive self-realization and the need to subordinate personal claims in order to obtain the co-operation of one's fellows.

RAYMOND FIRTH

Studies in Reading. Vol. I. Publications of the Scottish Council for Research in Education, no. xxvi. London: University of London Press. 1948. Pp. xv + 212. 5s. net.

This volume contains four main papers and the reports of three supplementary investigations concerned with the early stages of learning to read.

Dr W. B. Inglis writes a comprehensive and able review of recent work in this field. He covers eye movements, visual and auditory factors, laterality, reading disabilities and so on which are necessary for a basic understanding of the psychology of reading, but he stresses that reading is not 'barking at print' but a mental accomplishment concerned with getting meaning from the printed words and he devotes much attention to the influence of speech background and social factors on reading progress. An evaluation of methods of teaching reading and a discussion of reading readiness conclude a survey which will be found both a useful introduction to the subject and also a rich source of suggestions for further research. A bibliography of 93 titles is appended. Most of these are American—which is just, but the work of Burt, Schonell, M. D. Vernon, MacMeeken and Watts receives due notice.

In two valuable papers Dr P. E. Vernon reports the results of investigations into the speech vocabulary of Scottish children at the age of entering school and the contents of the primers and Infant Readers. The conversations of 200 children between the ages of $4\frac{1}{2}$ and $5\frac{1}{2}$ are analysed to discover the words they use and the topics they discuss. The 2900 different words noted are given in an alphabetical list with the frequency of their occurrence and the 491 commonest words are again classified by parts of speech. In spite of some defects in the manner of collecting this material the word counts probably give a reliable indication of the speech vocabulary and interests of children at this age. A word count of the Readers is then made and the results are compared with the speech vocabulary and an American Reading Vocabulary constructed mainly on the basis of Gates's list. The two word counts are analysed in great detail and Dr Vernon makes some illuminating comments on the many interesting facts which emerge. The main conclusions are that the speech vocabularies of Scottish and American children at this level are very similar, but there is too wide a discrepancy between the speech vocabularies and the Readers. Further, by comparison with the topics of the conversations, the content of the Readers is often remote from the interests and experience of the children.

These vocabularies receive further study in three supplementary investigations. Miss Carrigan sought

to discover how many of the words used in the Readers were known to the children. She found that the average 5-year-old knew about three-quarters but some less than 30 %, so that there is an obvious need to suit the reading material to the individual's stage of development. Dr McAllister compares the words in the Readers with the commoner words in the Speech Vocabulary and the 850 words of Basic English and Miss McCallien presents a phonetic analysis of the consonants in the words of these three lists.

The last paper is a most interesting study by Sister Jude of the vocabulary of Comics. She made a word count of the main coloured picture pages of five comics and the comic strip of a Sunday newspaper popular with 6- and 7-year-old children and compared her results with the Speech and Reader Vocabularies. She found that 74 % of the words were to be found in one or other of Vernon's lists, while of those not included the majority were words worth knowing. The irregularities of language amounted to no more than 6 % and seemed to have no harmful effect on the children's language. Representative word lists were constructed and administered as reading tests. Most of the words could be read by children of 6 and 7 and many of those not known in isolation could be read in their picture context. Comics are studied with intense interest and enjoyment by children; they stimulate the desire to read and the picture setting provides useful clues to aid the more backward in the learning of new words. It would seem that comic strips or material constructed on similar lines should have great usefulness in the early stages of reading and even more, perhaps, in remedial teaching.

Altogether this is a most valuable collection of studies on an important topic. We shall look for signs of its influence on the construction of new Infant Readers and shall await with interest the appearance of the promised volume II in this series.

A. H. D. TOZER

Aids to Educational Research. Publications of the Scottish Council for Research in Education, no. xxxi. London: University of London Press. 1949. Pp. viii + 32. 1s. net.

As a tool in the practical execution of research, parts of this pamphlet will be useful to post-graduate students working in general psychology as well as to the Scottish research workers in education for whom it is chiefly designed. It consists of lists of books on research methods and the presentation of findings, relevant periodicals, books on statistics, yearbooks, etc. Of less certain value is a list of topics awaiting investigation, classified according to importance and urgency. No doubt the selection of topics represents competent opinion at the present time, and it is to be hoped that the topics will be investigated. But it would have been well to recognize explicitly that the most valuable lines of research can very often not be foreseen by supervisors and administrators, still less by committees. The programme framed here suggests tacitly that research consists in finding answers to familiar questions; it ignores the importance of discovering questions that have not previously been asked.

D.W.H.

Psychology and Military Proficiency. By CHARLES W. BRAY. Princeton University Press (London: Geoffrey Cumberlege). 1948. Pp. xvii + 242. 20s. net.

This book, written by the chief of the Applied Psychology Panel of the U.S. National Defence Research Committee, provides an excellent general summary 'for all intelligent men' of the results achieved by psychologists working for that Panel. The style is clear and interesting and contains comparatively little technical detail. Nevertheless, very full references are provided so that psychologists will know where to look for more detailed information. It should, however, be borne in mind that Prof. Bray states that 'the war added little to methodology, but only proved again the practical value of existing techniques for the study of human nature'.

The book contains three chapters on the organization of psychological research work, and six chapters describing research projects, which were concerned mainly with classification, training and equipment, or combinations of the three.

The chapters on the organization of research make it clear that this was quite different from that which developed in this country during the last war. Teams from universities or research organizations set up research centres on service locations for specific projects. This had disadvantages which are freely acknowledged by Prof. Bray. First, civilian psychologists needed some months before they were sufficiently acclimatized to understand adequately the military context of their research work. Secondly, there were considerable organizational difficulties in carrying out classification procedures. Thirdly, the research findings were often ignored by the Services. This last deficiency is factually reported in connexion with several projects, and perhaps most clearly on page 187, where the following paragraph occurs: 'The

recording system for field artillery was completed in the summer of 1945 at the termination of the work of the project. The system was offered to Army Ground Forces for further and more complete experimentation on field artillery errors. Since the Army was not interested the Panel sold the materials for salvage'.

In spite of all these organizational difficulties, it is clear that much extremely valuable work was done, and one project on Radio Code Operators is recorded by Prof. Bray as 'a model study in military psychology, at least within the limited fields of selection and training'. Few readers will disagree with this description.

Turning to the research projects, the general conclusions are very like those reached by service psychologists in this country. Perhaps the greatest contribution of psychologists was in teaching the services that wars cannot be won either by man or by machine alone, but that 'man-machine' is the unit which must be studied and rendered efficient.

Another major lesson was that psychologists must study classification, training and equipment if good results are to be achieved. This finding led in turn to what is perhaps the most important fundamental principle in military psychological research. This is that the first task of the psychologist must always be to develop reliable measures of human performance. Only when these have been provided can research become really effective. This is perhaps best summarized on page 218 where Prof. Bray says 'The development of criterion measures, then, characterized the psychological approach to military problems. It was common to psychological studies and it distinguished the psychological approach from that of the personnel administrator on the one hand and the engineer on the other hand. The personnel man applied general principles and common sense; evaluation of results was empirical. The engineer set standards for his material and measured the material to insure that it met the standards; engineering measurements were limited to material. The psychologist developed standards for the man in relation to the material and measured man's performance to insure that it met the standards.' Those whose efforts to validate selection procedures have been largely vitiated by inadequate criteria will agree whole-heartedly with this point of view.

The general level of scientific work was undoubtedly very high, and a genuine effort was made to give technical advice only when adequate research had been carried out. Nevertheless, in wartime conditions this was not always possible, and Prof. Bray admits (page 39) that 'recommendations for action were often made before the complete evidence was obtained. Had it been otherwise, the work would have had little practical value in the emergency.'

Thus a training device for cargo handlers was installed without being validated (page 104). The classification and training procedures developed for the Amphibious Training Command produced only low validity coefficients (page 103). In a training device 'the subordinate unit tasks were presented in ways believed to be sufficiently realistic' (page 135). 'The speech interview for selecting telephone talks was never validated' (page 141).

The dangers of such 'hit and miss' methods were, however, keenly appreciated and, whenever possible, sounder methods devised. An interesting example arose in the training of aircrew to use the airplane interphone systems. Here a training system was devised by psychologists and speech experts; then 'Student pilots trained on the course were compared for intelligibility with untrained student pilots. The trained men were no better than the untrained men. The course was of no value' (page 143). Detailed research followed and led in the end to a great increase in intelligibility.

In similar ways, as far as time and conditions allowed, objective measurement replaced guesswork and opinion, whenever the services passed their problems to the psychologists. The story of the always useful and sometimes spectacular results makes fascinating reading.

At one or two points in the narrative, we come across unexpected statements. For example, on page 220 we find 'aptitude tests... test the knowledge or skills which most of our population have had an opportunity to acquire'—a definition which hardly harmonizes with what is known about the inherent nature of aptitudes. Again, on page 197, it is implied that a team-scoring device which indicated the time taken by a team of men to solve a fire-control problem might be 'basic to... future psychological research on teamwork and perhaps on leadership'. Most psychologists in this country who have been concerned with problems of leadership will consider such devices to be only slightly relevant to the broader problems in selecting and training leaders.

These, however, are only minor defects in a thoroughly admirable book, which should be read by all who are interested in the applications of psychological research methods to practical problems.

B. UNGERSON

Social Psychology. By RICHARD T. LAPIERE and PAUL R. FARNSWORTH. Third Edition. London: McGraw-Hill. 1949. Pp. xiv+626. 27s. net.

Like many text-books on the borderline of sociology and psychology, this one is largely descriptive of the topics and materials with which a science of social psychology would deal, but it leaves the reader unconvinced that the science has established itself. However, it has the great merit of giving full references—brought up to date in this third edition—on most of the topics that concern the social psychologist, even when the work referred to is not assimilated into the authors' handling of the topic. With this fulness of reference in the notes goes an unfortunate verbosity in the text. Nothing is taken for granted, least of all any quickness of comprehension in the reader. Obvious generalizations are rammed home with banal instances, laborious similes are developed to elucidate elementary ideas, and tediously long illustrations of familiar facts are drawn from hypothetical cases.

The new edition retains the broad plan of earlier editions. Roughly the first third of the book deals with the social training of the individual, giving more attention to the function of expressive action (or 'gesture') and of language than is usual. A section on personality includes both a discussion of psychological theories and full illustration of the fact that various social groupings influence the personalities of their members and exercise some control over deviants. There follows a section dealing with the 'adjustments' (e.g. social, occupational, marital) that the individual has to make and the maladjustments that may occur (e.g. criminal and psychopathological). The final section, headed 'Social Interaction', discusses collective behaviour and treats of institutions, publics, mobs, propaganda, rumour, etc. The detailed attention given to various types of public and audience is a useful feature of the discussion, and here especially the references to other studies and descriptive sources are ample and valuable.

The general trend of the treatment is to overemphasize the influence of society on the individual. In their recapitulation the authors say 'The theme of this book is that the individual plays but a small role on a great and crowded stage and that he plays that role largely in accordance with a socially predetermined script. . . . Should he, furthermore, by force of social circumstances be one of those few who are equipped and expected to provide some measure of individual leadership for others, success will depend mainly on his happening to lead them in the direction in which they happen to be going.' In this onesidedness the authors undoubtedly reveal the influence of a particular time and place on their own thinking, but not the outcome of a science. They are throughout prone to accept fashions and prejudices in psychological concepts, without any convincing demonstration of their value. Thus 'stereotypes' are laughed out of court, but the sound and simple observations that underlay the moderate use of such generalizations by laymen are brought back—with a flourish of new achievement—under the heading of 'social, or group, personality norms'. Instinct theories are complacently derided, but no alternative answer is given to the problems they were meant to solve; even McDougall knew of social influences but they were not the answer to the questions he was asking. The same complacency in knocking down men of straw appears in the dismissal of psycho-analysis as a faith system like Christian Science. The general effect is to make the book seem parochial in outlook. The authors say of Kardiner, Linton and Kluckhohn: 'So far, their labors have consisted mainly of bringing to the attention of anthropologists, usually in new and confusing terminology, the more commonplace of the established concepts of modern social psychology.' This may be so, but it is a dangerous reflexion in a book which, often in a new and confusing terminology, calls our attention so largely to commonplaces of ordinary educated observation.

D.W.H.

Crime and Abnormality. By CECIL BINNEY. London: Oxford University Press. 1949. Pp. viii+176. 5s. net.

The title 'Crime and Abnormality' may be misleading to psychologists but the book will be welcome. It is a lucid discussion by a barrister of judicial problems arising from the law on crime and abnormality. As such it could hardly be bettered for the general reader. Acutely reasoned and exact in statement it is so lightened by illustrative instances as to be steadily interesting and readable. A careful account of English procedure in the trial of persons who may be insane is followed by a long section on 'Insanity as a Defence' and shorter chapters on 'Mental Defectives', 'Drunkenness and Crime' and 'Infanticide and Suicide'. A final chapter discusses the broader questions of justice and sentiment that affect the practical application of the law. The author refers frankly to conflict between lawyers and doctors and he writes from the lawyer's standpoint, but psychologists who disagree with some of his opinions will still be grateful for so scrupulous and reasonable a statement of the framework which supports legal thinking on this topic.

Thematic Apperception Test: Thompson Modification. By CHARLES E. THOMPSON. Cambridge, Mass.: Harvard University Press (London: Geoffrey Cumberlege). 1949. 30 plates and Manual. 27s. 6d. net.

Negro Americans were found to produce meagre and sometimes stereotyped material in response to the ordinary T.A.T. pictures, apparently because of barriers to their identification with the white people depicted. This difficulty has been dealt with in the Thompson Modification by having most of the T.A.T. pictures (1943 edition) redrawn with Negro faces and dark skin colouring for all the figures. The author reports that the stories given by Negroes for the modified pictures 'showed an empathy which is comparable to the empathy shown by the white subjects telling stories for the T.A.T.' He notes a qualitative impression that with the revised pictures Negroes also express more bitterness about discriminatory social practices and give the characters more violent and destructive hatred than they did when responding to the ordinary T.A.T. pictures.

Modern Discoveries in Medical Psychology. By CLIFFORD ALLEN. Second edition. London: Macmillan. 1949. Pp. xii + 236. 12s. 6d. net.

This revised edition of a popular book intended for general practitioners, medical students and the intelligent layman includes a new chapter on physical methods of treatment. But the emphasis of the book remains on the approach to mental disorder through an understanding of the mind, for the author believes that an eventual appraisal of organic and physical methods may lead to a new swing of the pendulum towards psychotherapy. It is surprising that in a second edition the chapter on 'Kretschmer and the relation of body and mind' should make no mention of Sheldon's work. The fact is that the 'Modern Discoveries' of the title is by now misleading; the book really offers an introduction to the work of pioneers in modern medical psychology.

OTHER PUBLICATIONS RECEIVED

- ADLER, ALFRED. *Menschenkenntnis*. Zurich. 1947. Pp. vii+236.
- BAUMGARTEN, F. *Orientation et Selection Professionnelles par l'Examen Psychologique du Caractère*. Paris: Dunod. 1949. Pp. vii+183.
- BAUMGARTEN, F. *Psychologie et facteur humain dans l'entreprise*. Neuchatel and Paris: Delachaux et Niestle. 1948. Pp. 196.
- BIERER, JOSHUA (ed.). *Therapeutic Social Clubs*. London: H. K. Lewis. N.D. [1948]. Pp. 74. 5s. net.
- BRADFORD, ELIZABETH. *Let's Talk about Children*. Kingswood: The World's Work. 1949. Pp. x+134. 7s. 6d. net.
- BRINK, FREDERICK W. *This Man and This Woman*. Kingswood: The World's Work. 1949. Pp. viii+76. 5s. net.
- COLKET-CANER, G. *It's How You Take It*. Kingswood: The World's Work. 1948. Pp. viii+156. 8s. 6d.
- DAS, ADHAR CHANDRA. *Negative Fact, Negation and Truth*. University of Calcutta. 1942. Pp. xxiii+296.
- FAITHFULL, THEODORE. *Letters to Margaret*. London and Birmingham: Rylee Ltd. 1948. Pp. 108. 5s. 6d. net.
- FAITHFULL, THEODORE. *A Handbook of Self Analysis*. London and Birmingham: Rylee Ltd. 1948. Pp. 116. 5s. 6d. net.
- FRANKAU, GILBERT. *Michael's Wife*. Macdonald. 1948. Pp. 546. 12s. 6d. net.
- GERMANSKY, B. *Psychologische Kategorialistik, Eine Neue Wissenschaft*. Jerusalem: Hamadpis Liphshitz Press. 1949.
- IKIN, A. GRAHAM. *Religion and Psychotherapy*. London and Birmingham: Rylee Ltd. 1948. Pp. 112. 5s. 6d. net.
- Interim Report—Children's Reception Centre, Mersham, Kent*. London: Headley Bros. March 1949. Pp. 35. 3d. net.
- INTERNATIONAL COMMITTEE OF THE RED CROSS, MEDICAL DIVISION. *Health Conditions among the Civilian Populations of Certain Countries Affected by the War*. Geneva. 1949.
- KISS, P. *Az Ember Felé*. Budapest: Egyetem Nyomda. 1948. Pp. 236.
- MACAULAY, MARY. *Understanding Ourselves*. London: Routledge and Kegan Paul. 1949. Pp. v+148. 7s. 6d. net.
- MCQUEEN, J. A. *Intelligence Work and the Bible*. London: Covenant Publishing Company. 1949. Pp. vi+103. 2s. 6d. net.
- MENNINGER, W. C. *Psychiatry—Its evolution and Present Status*. New York: Cornell University Press; London: Oxford University Press. 1949. Pp. xiv+138. 11s. 6d. net.

This book, already reviewed in this *Journal* for March 1949, has now been issued by an English publisher.

MULLAHY, PATRICK. *Oedipus, Myth and Complex*. London: Allen and Unwin. 1949. Pp. xxii+538. 37s. 6d. net.

This book, already reviewed in this *Journal* for September 1949, has now been issued by an English publisher.

- SORENSEN, H. & MALM, MARGUERITE. *Psychology for Living*. London: McGraw Hill. 1948. Pp. x+638. 18s. net.
- STERNEGGER, BENEDIKT. *Geheimnisse der menschlichen Seele*. Augsburg: Manu Verlag. 1948. Pp. 207. DM. 4.90.
- WALL, C. WILFRED. *The Value of Neurosis*. London and Birmingham: Rylee Ltd. 1948. Pp. 116. 5s. 6d. net.
- WILLIAMSON, E. G. *Trends in Student Personnel Work*. Minnesota University Press; London: Geoffrey Cumberlege. 1949. Pp. viii+417. 27s. 6d. net.